Efficient Non-Local Dynamical Electron Correlation for Condensed Matter Simulations

---

Dissertation

zur

Erlangung der naturwissenschaftlichen Doktorwürde
(Dr. sc. nat.)

vorgelegt der

Mathematisch-naturwissenschaftlichen Fakultät

der

Universität Zürich

von

Mauro Del Ben

von

Italien

Promotionskomitee

Prof. Dr. Jürg Hutter (Vorsitz)

Prof. Dr. Joost VandeVondele (Leitung der Dissertation)

Zürich, 2015

Abstract

Density Functional Theory (DFT) has become the most widely used quantum mechanical tool in chemistry and physics for predicting properties of materials ranging from molecules to condensed phase systems. However, current approximate DFT still suffers from many flaws and the quest for finding more accurate functional is still ongoing. In this respect, the second-order Møller-Plesset perturbation theory (MP2) and the Random Phase Approximation (RPA), are increasingly popular post-Kohn-Sham electronic structure methods. Both MP2 and RPA display many appealing features such as the capability to describe covalent, ionic, hydrogen-bond and dispersion interactions accurately and from first principles. On the other hand, these advantages come at a computational cost that is significantly higher than that of traditional DFT. On this basis, the development of efficient algorithms, capable to exploit the performance of large supercomputers facilities, is of prime interest in order to extend the applicability of these methods to larger and more realistic systems. The main achievement of this work is the development of a novel algorithm for MP2 and RPA correlation energies of finite and extended systems based on a hybrid Gaussian and Plane Waves (GPW) approach with the resolution-of-identity approximation (RI). The key aspect of the method relies in the dual representation of the RI fitting densities in term of Gaussian and Plane Waves auxiliary functions, leading to a simplified treatment of the Coulomb interactions that is particularly efficient in the condensed phase. The RI approximation allows to speed up the MP2 energy calculations by a factor 10 to 15 compared to the canonical implementation, but still requiring $O(N^5)$ operations. On the other hand, in the RPA case, the combination of RI and imaginary frequency integration reduces the computational effort from $O(N^6)$ to $O(N^4 \log N)$. Our implementation has low memory requirements and displays excellent parallel scalability up to several thousands of processes. Furthermore, the computationally more demanding parts, that is the $O(N^5)$ and $O(N^4)$ scaling steps respectively for MP2 and RPA, can be accelerated by employing graphics processing units (GPU) showing, for large systems, a gain of more than a factor three compared to the standard only CPU case. In this way, RI-MP2 and RI-RPA calculations for condensed phase systems containing hundreds of atoms and thousands of basis functions can be performed within minutes employing few hundreds hybrid nodes. Moreover, at the MP2 level of theory, the equations for calculating the forces acting on atoms and the stress tensor have been derived and implemented in the framework of the RI-GPW method. The developed algorithm, suitable for massively parallel computation, displays timing for evaluating the forces and stresses similar to calculating the energy only. This allowed to use these high level electronic structure methods for performing first principle simulations on a variety of systems with a particular focus on the study of the structural properties of bulk liquid water under ambient conditions and ice at high pressure.

Zusammenfassung

Dichtefunktionaltheorie (DFT) ist ein weit verbreitetes Verfahren, um chemische und physikalische Fragestellungen auf quantenmechanischer Basis zu beantworten. DFT wird verwendet um Eigenschaften von Molekülen und Festkörpern zu berechnen. Trotz des grossen Erfolgs dieser Methode sind die in DFT angenommen Nährungen teilweise zu grob und führen in bestimmen Fällen zu grossen Fehlern. Die Suche nach besseren Funktionalen ist auch heute noch ein intensiv bearbeitetes Forschungsgebiet. Aus diesem Grund erfreuen sich akkuratere Methoden wie Møller-Plesset Störungstheorie (MP2) und die sogenannte "Random Phase Approximation" (RPA) einer zunehmenden Beliebtheit. Mit MP2 und RPA lassen sich sowohl kovalente, ionische, und dispersive Wechselwirkungen als auch Wasserstoffbrüeckenbindungen sehr genau *ab initio* bestimmen. Der Nachteil im Vergleich zu DFT ist, dass die Rechnungen wesentlich aufwändiger sind und signifikant mehr Rechnenzeit bei gleich bleibender Systemgroesse benötigen. Die Zielsetzung dieser Arbeit ist die Entwicklung von neuen Algorithmen zur Berechnung von MP2 und RPA Korrelationsenergien für nicht-periodische und periodische Systeme, basierend auf der "Gaussian and Plane Waves (GPW)"- und "resolution-of-identity (RI)"- Näherung. Der Kern der Methode ist die duale Repräsentation der gefitten RI-Dichte mit Hilfe von Gaussfunktionen und ebenen Wellen als Hilfsfunktionen. Dies lässt eine einfachere und effizientere Berechnung des Coloumbpotentials insbesondere für periodische Systeme wie Festkörper zu. Verglichen mit der kanonischen Implementierung beschleunigt die RI-Näherung die Berechnung der MP2-Energien um das 10- bis 15-Fache. Dennoch werden $O(N^5)$ Rechenoperationen benötigt. In RPA wird die RI-Nährung mit der Integration von imaginären Frequenzen kombiniert, was die die Ordnung der Rechnung von $O(N^6)$ auf $O(N^4 \log N)$ senkt. Unsere Implementierung ist hinsichtlich des Arbeitsspeicherverbrauchs effizient und skaliert exzellent auf bis zu 10 000 parallele Prozesse. Weiterhin werden graphics processing units (GPUs) genutzt, um die $O(N^5)$ und $O(N^4)$ Rechenschritte für MP2 bzw. RPA zu beschleunigen. Für grosse Systeme lässt sich die Rechenzeit damit um den Faktor 3 reduzieren. Damit ist es möglich, periodische Systeme mit hunderten Atomen und tausenden Basisfunktionen innerhalb von Minuten auf einigen hundert CPU-GPU Hybridcomputern mit RI-MP2 und RI-RPA zu berechnen. Für MP2 sind die Kräfte, die auf die Atome wirken, und der Spannungstensor hergeleitet und im Rahmen der RI-GPW-Methode implementiert worden. Die Berechnung der MP2-Kräfte wurde algorithmisch für massive Parallelisierung optimiert. Mit dieser Implementierung ist die Berechnung der Kräfte und Spannungstensoren auf dergleichen Zeitskala wie die der MP2-Energien möglich. Diese akkuraten *ab initio* Methoden zur Bestimmung der Elektronenstruktur wurden für eine breite Spannweite von Systemen verwendet. Intensiv wurde insbesondere flüssiges Wasser unter Standardbedinüngen und Eis bei hohem Druck untersucht.

Acknowledgements

I would like to thank Prof. Dr. Jürg Hutter and Prof. Dr. Joost VandeVondele for giving me the opportunity to do my PhD. This project would not have been possible without their help and advice. The developments and most of the calculations reported here were enabled by the Swiss National Supercomputer Centre (CSCS) under project IDs s441 and ch5. The research leading to these results has received funding from the Swiss University Conference through the High Performance and High Productivity Computing (HP2C) programme and the Platform for Advanced Scientific Computing (PASC) programme. This project has also used resources of the Oak Ridge Leadership Computing Facility at the Oak Ridge National Laboratory, which is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC05-00OR22725, under a DD project ID MAT106. Additionally, intense production calculations have been achieved using the PRACE (Partnership for Advanced Computing in Europe) Research Infrastructure resource Hermit based in Germany at Stuttgart (HLRS). Finally, I would like to express my gratitude to all people that shared with me the ups and downs of life as a PhD student.

Publications

1. Mauro Del Ben, Jürg Hutter, and Joost VandeVondele "Force and Stress in Second Order Møller-Plesset Perturbation Theory for Condensed Phase Systems within the Resolution-of-Identity Gaussian and Plane Waves Approach" *J. Chem. Phys.* Accepted (2015).

2. Mauro Del Ben, Ole Schütt, Tim Wentz, Peter Messmer, Jürg Hutter, and Joost VandeVondele "Enabling Simulation at the Fifth Rung of DFT: Large Scale RPA Calculations with Excellent Time to Solution" *Comput. Phys. Commun.* **187**, 120 (2015).

3. Mauro Del Ben, Joost VandeVondele, and Ben Slater "A Periodic MP2, RPA and Boundary Condition Assessment of Hydrogen Ordering in Ice XV" *J. Phys. Chem. Lett.* **5**, 4122 (2014).

4. Mauro Del Ben, Mandes Schönherr, Jürg Hutter, and Joost VandeVondele "Correction to Bulk Liquid Water at Ambient Temperature and Pressure from MP2 Theory" *J. Phys. Chem. Lett.* **5**, 3066 (2014).

5. Mauro Del Ben, Mandes Schönherr, Jürg Hutter, and Joost VandeVondele "Bulk Liquid Water at Ambient Temperature and Pressure from MP2 Theory" *J. Phys. Chem. Lett.* **4**, 3753 (2013).

6. Mauro Del Ben, Jürg Hutter, and Joost VandeVondele "Electron Correlation in the Condensed Phase from a Resolution of Identity Approach Based on the Gaussian and Plane Waves Scheme" *J. Chem. Theory Comput.* **9**, 2654 (2013).

7. Mauro Del Ben, Jürg Hutter, and Joost VandeVondele "Second-Order Møller-Plesset Perturbation Theory in the Condensed Phase: An Efficient and Massively Parallel Gaussian and Plane Waves Approach" *J. Chem. Theory Comput.* **8**, 4177 (2012).

8. Mauro Del Ben, Remco W. A. Havenith, Ria Broer, and Mauro Stener "Density Functional Study on the Morphology and Photoabsorption of CdSe Nanoclusters" *J. Phys. Chem. C* **115**, 16782 (2011).

Contents

Chapter 1

Introduction

The advent of quantum mechanics had a great impact in chemistry. In fact, quantum effects are dominating the phenomena at the atomic and molecular scale, that represents the size regime of interest to chemists. The main achievement was that it was possible to write down the equation providing the exact treatment of any quantum system, namely the Schrödinger equation (SE). Mathematically speaking the SE is a differential equation that once solved leads to a function, the wavefunction, that contains all necessary information for the description of a system. Additionally the form of the exact wavefunction, as well as a prescription for its calculation, was also known. The drawback is that the solution of the SE following this procedure results into an amount of data that is practically impossible to handle and approximations need to be considered. A series of approaches were thus introduced, allowing for a systematic improvement of the quantum description of a system, giving birth to wavefunction theory, the foundation of modern quantum chemistry. Despite this, the inherent high computational cost associated even with the lower level wavefunction based methods, limited for a long time the widespread application of these approaches.

The breakthrough came with the introduction of density functional theory within the Kohn-Sham formulation (KS-DFT). The reason why KS-DFT was successful is related to the usage of electronic density as fundamental variable instead of the wavefunction. In fact the former is a much simpler object than the latter since it depends only on three coordinates in space instead of explicitly treating all particles simultaneously, and can thus be handled even with limited computational resources. Moreover, DFT provides a formally exact way to solve the Schrödinger equation [1, 2] and thus it offers the appealing possibility to treat all the underling physics of a system in a quantum mechanical and rigorous way. Unfortunately this heavenly picture faces the problem that the exact treatment of exchange and correlation effects is unknown and practical DFT calculations account for them by using one of the various models, referred to as density functional approximations (DFA). Tremendous improvements have been achieved in the direction of increasing the accuracy of DFA, and Perdew *et al.* [3] classified them on a "Jacob's ladder" for which each rung of the ladder introduces more descriptors of the electronic system and yields models with improved accuracy. At the present stage, this ladder has five rungs making use progressively of the information on 1) the electronic density, 2) its gradient, 3) the kinetic energy density, 4) the occupied molecular orbitals (MO) and 5) the virtual MO's. At the fifth rung the inclusion of the virtual orbitals allows for taking into account

the non-local dynamical electron correlation contributions, responsible, among these, of the long-range van der Waals (vdW) dispersion interactions. The dynamical electron correlation is a purely quantum mechanical effect and can be considered as a consequence of the fact that electrons will repel each other as they move around. The entangled (correlated) motion of the electrons results in the tendency of decreasing the repulsion energy, and has important implications in many chemical and physical phenomena.

Many of the various functionals on the 5^{th} rung are based on either the random phase approximation (RPA) [4–12], or many body perturbation theory (MBPT) in the form of double hybrids (DH) [13–19]. Even though these methods represent well known approaches since the early time of quantum mechanics, their introduction in the context of DFT is relatively recent. The reasons for this can be found either in the ongoing quest for fixing known flaws of DFT, needed for treating complex systems, and in the significantly higher computational cost associated with these methods. In this respect the development of low order scaling algorithms as well as efficient implementations, capable to exploit state-of-the-art supercomputer facilities, is of great importance in order to increase the applicability of these techniques.

The central objective of this work is the development and application of such methodologies with the focus on condensed phase systems. The thesis is organized in nine chapters. After this short introduction, Chapter 2 briefly presents the theoretical framework of the basis of many body perturbation theory, in the formulation of Møller-Plesset, and random phase approximation. Between Chapter 3 and Chapter 6 the publications related to the method development and high performance computer implementation are summarized. In particular, Chapter 3 reports the basis of the Gaussian and Plane-Waves (GPW) scheme applied to the calculation of the second order Møller-Plesset correlation energy (MP2). The GPW method, originally developed for handling the ground state electrostatic density/potential, is extended for dealing with occupied-virtual densities, giving a general procedure for treating correlation energies in the condensed phase. In Chapter 4 the resolution of identity (RI) approximation is introduced and combined with the GPW scheme. The RI approximation allows to speed up the MP2 energy calculations by a factor 10 to 15 compared to the canonical implementation, but still requiring $O(N^5)$ operations, with N proportional to the system size. On the other hand, in the RPA case, the combination of RI and imaginary frequency integration reduces the computational effort from $O(N^6)$ to $O(N^4 \log N)$. The focus of Chapter 5 is on high performance computer implementation of the RPA method, with the target of making this approach feasible for very large systems. Finally Chapter 6 presents the equations for evaluating the forces acting on the atoms and the stress tensor of the MP2 energy in the framework of the RI-GPW method. Based on these a massively parallel algorithm has been developed for finite and extended systems. Applications are presented in Chapter 7 and Chapter 8. The former deals with bulk liquid water, representing one of the most challenging condensed phase systems for computer simulation. This system has been studied at the MP2 level, the first reported simulation with this approach. The latter reports an MP2 and RPA study on the structural stability and properties of ice XV. Chapter 9 summarizes the main achievements and gives an overview for further developments.

Chapter 2

Theory

In this chapter, the theoretical framework employed in the following chapters is concisely presented. The aim is to give a general introduction on the methods that represent the basis of the approaches that have been developed over this thesis work, namely electron correlation energy from many body perturbation theory and the random phase approximation. In particular the Hartree-Fock approximation and Density Functional Theory, representing the foundations of modern quantum chemistry and computational material science, are described in general terms. More details of the topics can be found in standard textbooks [14, 20–23] or in the review papers cited in the different sections of this chapter.

2.1 Fundamentals

2.1.1 Born-Oppenheimer Approximation

One of the central goals of modern quantum chemistry is to find approximated solutions to the non-relativistic time-independent Schrödinger equation:

$$H|\Psi\rangle = E|\Psi\rangle \tag{2.1}$$

where H is the Hamilton operator of a system made of M nuclei and N electrons and E the total energy. Given that the positions of the nuclei and electrons with respect to a fixed coordinate system are described respectively by vectors $\{\vec{R}_A\}$ and $\{\vec{r}_i\}$, the H operator takes, in atomic units, the following form:

$$\begin{aligned} H = &-\frac{1}{2}\sum_{i}^{N}\nabla_i^2 - \frac{1}{2}\sum_{A}^{M}\frac{1}{M_A}\nabla_A^2 - \sum_{i}^{N}\sum_{A}^{M}\frac{Z_A}{|\vec{r}_i - \vec{R}_A|} \\ &+ \sum_{i<j}^{N}\frac{1}{|\vec{r}_i - \vec{r}_j|} + \sum_{A<B}^{M}\frac{Z_A Z_B}{|\vec{R}_A - \vec{R}_B|} \end{aligned} \tag{2.2}$$

where M_A and Z_A are respectively the mass and charge of the nucleus A. The terms in Equation (2.2) represents respectively the kinetic energy of the electrons, the kinetic energy of the nuclei, the Coulomb interaction between electrons and nuclei, the Coulomb repulsions between electrons and the Coulomb repulsions between nuclei. The kinetic

energy operator (the Laplacian ∇^2), involves the calculations of the derivatives with respect to the coordinates of each particle. In Equation (2.2) the short hand notation $i < j$ and $A < B$ in the Coulomb repulsions terms means that the double sum run only over unique pairs.

Therefor, the exact solution of Equation (2.1) implies that the wavefunction (WF) Ψ depends simultaneously on the coordinates of the electrons and nuclei. It is in fact not possible to factorize Ψ as a product of an electronic and nuclear WF due to the third term in Equation (2.2) that couples the different kind of particle coordinates. To overcome this problem it can be argued that, due to the fact that electrons are much lighter than nuclei, the former are moving into a fixed frame of the latter, adjusting their positions on-the-fly every time the nuclei are moving. According to this assumption, that is the $\{\vec{R}_A\}$ are fixed, the kinetic energy of the nuclei can be neglected, while the Coulomb repulsions between nuclei becomes a constant. This allows to define the electronic Hamilton H_{elec} that depends explicitly on the electronic degrees of freedom and parametrically on the nuclear coordinates:

$$H_{\text{elec}} = -\frac{1}{2}\sum_{i}^{N}\nabla_i^2 - \sum_{i}^{N}\sum_{A}^{M}\frac{Z_A}{|\vec{r}_i - \vec{R}_A|} + \sum_{i<j}^{N}\frac{1}{|\vec{r}_i - \vec{r}_j|} \tag{2.3}$$

The solution of the Schrödinger equation employing the H_{elec} delivers the electronic WF Ψ_{elec} and associated energy E_{elec} for the considered configuration of nuclei. The total energy in this case is obtained simply by adding to E_{elec} the constant term derived from Coulomb repulsions between nuclei

$$E_{\text{tot}}(\{\vec{R}_A\}) = E_{\text{elec}} + \sum_{A<B}^{M}\frac{Z_A Z_B}{|\vec{R}_A - \vec{R}_B|} \tag{2.4}$$

where the parametric dependence of E_{tot} on $\{\vec{R}_A\}$ has been emphasized.

By inspection of Equation (2.3) and Equation (2.4) it can be noted that $E_{\text{tot}}(\{\vec{R}_A\})$ contains all term in Equation (2.2) except for the kinetic energy of the nuclei. It is thus possible to define the nuclear Hamiltonian H_{nuc} as

$$H_{\text{nuc}} = -\frac{1}{2}\sum_{A}^{M}\frac{1}{M_A}\nabla_A^2 + E_{\text{tot}}(\{\vec{R}_A\}) \tag{2.5}$$

where $E_{\text{tot}}(\{\vec{R}_A\})$ provides the potential energy surface over which the nuclei are moving. The solution of the Schrödinger equation employing the H_{nuc} gives the Born-Oppenheimer (BO) approximation [24] to the total energy of Equation (2.1).

From now on, only the solution of the electronic Hamiltonian under the BO approximation will be considered and the subscript "elec" will be dropped.

2.1.2 The Variational Principle

Referring to Equation (2.1), let's assume that there exists a set of exact solutions, infinite in number

$$H|\Psi_i\rangle = E_i|\Psi_i\rangle \qquad i = 0, 1, \ldots \tag{2.6}$$

where

$$E_0 \leq E_1 \cdots E_i \leq \cdots \tag{2.7}$$

Since the Hamiltonian operator is Hermitian, the eigenvalues $\{E_i\}$ are real and the relative eigenfunctions $\{\Psi_i\}$ are orthonormal, *e.g.* $\langle\Psi_j|\Psi_i\rangle = \delta_{ij}$, with δ_{ij} being the Kronecker delta. In this respect Equation (2.6) can be rewritten in an equivalent form as:

$$\langle\Psi_j|H|\Psi_i\rangle = E_i\delta_{ij} \tag{2.8}$$

Additionally, $\{\Psi_i\}$ create a complete basis, that is any function $\tilde{\Phi}$, that fulfills the same boundary conditions, can be expressed as a liner combinations of Ψ_i's

$$|\tilde{\Phi}\rangle = \sum_i |\Psi_i\rangle c_i = \sum_i |\Psi_i\rangle\langle\Psi_i|\tilde{\Phi}\rangle \tag{2.9}$$

The variational principle states that for any normalized trial wavefunction $|\tilde{\Phi}\rangle$ the associated expectation value of the Hamiltonian is an upper bound of the exact ground state energy E_0, that is:

$$\langle\tilde{\Phi}|H|\tilde{\Phi}\rangle \geq E_0 \tag{2.10}$$

$$\langle\tilde{\Phi}|\tilde{\Phi}\rangle = 1 \tag{2.11}$$

where the equality is fulfilled if $|\tilde{\Phi}\rangle = |\Psi_0\rangle$. This inequality is proved by first inserting Equation (2.9) into the normality conditions, Equation (2.11), and exploiting the orthonormality of the $\{\Psi_i\}$:

$$\begin{aligned}\langle\tilde{\Phi}|\tilde{\Phi}\rangle &= \sum_{ij}\langle\tilde{\Phi}|\Psi_i\rangle\langle\Psi_i|\Psi_j\rangle\langle\Psi_j|\tilde{\Phi}\rangle = \sum_{ij}\langle\tilde{\Phi}|\Psi_i\rangle\delta_{ij}\langle\Psi_j|\tilde{\Phi}\rangle \\ &= \sum_i\langle\tilde{\Phi}|\Psi_i\rangle\langle\Psi_i|\tilde{\Phi}\rangle = \sum_i \left|\langle\Psi_i|\tilde{\Phi}\rangle\right|^2 = 1\end{aligned} \tag{2.12}$$

the same procedure applied to $\langle\tilde{\Phi}|H|\tilde{\Phi}\rangle$ leads to:

$$\langle\tilde{\Phi}|H|\tilde{\Phi}\rangle = \sum_{ij}\langle\tilde{\Phi}|\Psi_i\rangle\langle\Psi_i|H|\Psi_j\rangle\langle\Psi_j|\tilde{\Phi}\rangle = \sum_i E_i \left|\langle\Psi_i|\tilde{\Phi}\rangle\right|^2 \tag{2.13}$$

Since $E_i \geq E_0$ for all i it follows that

$$\langle\tilde{\Phi}|H|\tilde{\Phi}\rangle \geq \sum_i E_0 \left|\langle\Psi_i|\tilde{\Phi}\rangle\right|^2 = E_0 \sum_i \left|\langle\Psi_i|\tilde{\Phi}\rangle\right|^2 = E_0 \tag{2.14}$$

where in the last step the normality condition Equation (2.12) has been used.

The variational principle provides the criterion used when seeking the best possible solution for a given trial wavefunction $\tilde{\Phi}$. In fact, given a normalized $\tilde{\Phi}$, depending on a certain set of parameters, then the best solution can be found by varying these parameters till $\langle\tilde{\Phi}|H|\tilde{\Phi}\rangle$ reaches a minimum. This is the basis of the variational method.

2.1.3 Spin Orbitals and Slater Determinants

In general terms, in quantum chemistry, the single electronic wavefunctions are named *orbitals*. In order to fully describe an electron, an orbital function has to take the spatial distribution as well as the spin of the electron into account.

The former is defined by the spatial part of the orbital $\psi(\vec{r})$, that is solely a function of the electron position $\vec{r}$. The property of $\psi(\vec{r})$ is that its squared module gives the probability density $|\psi(\vec{r})|^2 \, d\vec{r}$ of finding the electron in the infinitesimal volume element $d\vec{r}$ at the position $\vec{r}$. This implies that spatial orbitals are normalized and generally they are also chosen to be orthogonal

$$\int \psi_p(\vec{r})^* \psi_q(\vec{r}) d\vec{r} = \delta_{pq} \tag{2.15}$$

In this respect, a complete set of spatial orbitals $\{\psi_p(\vec{r})\}$ can be used as a basis for expanding exactly any single particle wavefunction $\phi(\vec{r})$

$$\phi(\vec{r}) = \sum_p c_p \psi_p(\vec{r}) \tag{2.16}$$

with $\{c_p\}$ being constant values. In practice the number of spatial orbitals is a finite set $\{\psi_p(\vec{r}) \,|\, p = 1, \dots K\}$ that ultimately results into an error in the expansion of Equation (2.16) due to the basis incompleteness.

The spin part of the orbital is described by two orthonormal functions namely $\alpha(\omega)$ and $\beta(\omega)$, respectively referring to spin up ($\uparrow$) and down ($\downarrow$). The form of an orbital taking into account both the spatial distribution and spin can be obtained as the product of a spatial orbital with one of the spin functions. These single electron wavefunctions are named *spin orbitals* $\chi(\vec{x})$ for which $\vec{x}$ includes both the space ($\vec{r}$) and spin (ω) coordinates. Given a set of K spatial orbitals $\{\psi_p(\vec{r})\}$ it is thus possible to form $2K$ spin orbitals according to

$$\chi_q(\vec{x}) = \begin{cases} \psi_p(\vec{r})\alpha(\omega) \\ \quad \text{or} \\ \psi_p(\vec{r})\beta(\omega) \end{cases} \tag{2.17}$$

When moving from single-electron wavefunctions to a many-electron wavefunctions there are additional requirements that have to be taken into account. Given a collection of N electrons, the N-electron wavefunction Ψ^N depends on all electronic coordinates $\{\vec{x}_i \,|\, i = 1, \dots, N\}$, *i.e.* $\Psi^N = \Psi(\vec{x}_1, \dots, \vec{x}_N)$. In particular, since electrons are fermionic and indistinguishable particles, the antisymmetry principle requires that Ψ^N is antisymmetric (changes sign) with respect to the interchange of the space and spin coordinates of any two electrons

$$\Psi(\vec{x}_1, \dots, \vec{x}_p, \dots, \vec{x}_q, \dots, \vec{x}_N) = -\Psi(\vec{x}_1, \dots, \vec{x}_q, \dots, \vec{x}_p, \dots, \vec{x}_N) \quad \forall\, p \neq q \tag{2.18}$$

A function that fulfill these requirements can be formed from the evaluation of the following determinants

$$\Psi(\vec{x}_1, \dots, \vec{x}_N) = |\chi_1, \chi_2, \cdots \chi_N\rangle = \frac{1}{\sqrt{N!}} \begin{vmatrix} \chi_1(\vec{x_1}) & \chi_2(\vec{x_1}) & \cdots & \chi_N(\vec{x_1}) \\ \chi_1(\vec{x_2}) & \chi_2(\vec{x_2}) & \cdots & \chi_N(\vec{x_2}) \\ \vdots & \vdots & \ddots & \vdots \\ \chi_1(\vec{x_N}) & \chi_2(\vec{x_N}) & \cdots & \chi_N(\vec{x_N}) \end{vmatrix} \tag{2.19}$$

called Slater determinant (SD). The SD in Equation (2.19) has N electrons occupying N spin orbitals $\{\chi_p\}$. The Slater determinants have the following properties:

Table 2.1: Matrix elements between N-electron Slater Determinants for one-electron (O_1) and two-electron (O_2) operators [25, 26].

		$O_1 = \sum_i^N h(i)$	$O_2 = \sum_{i<j}^N \frac{1}{\lvert\vec{r}_i - \vec{r}_j\rvert}$
		$\langle\Psi\lvert O_1\rvert\Phi\rangle$	$\langle\Psi\lvert O_2\rvert\Phi\rangle$
Case 1	$\lvert\Psi\rangle = \lvert\cdots pq\cdots\rangle$ $\lvert\Phi\rangle = \lvert\cdots pq\cdots\rangle$	$\sum_p^N (p\lvert h\rvert p)$	$\frac{1}{2}\sum_p^N\sum_q^N \big[(pp\lvert qq) - (pq\lvert pq)\big]$
Case 2	$\lvert\Psi\rangle = \lvert\cdots pq\cdots\rangle$ $\lvert\Phi\rangle = \lvert\cdots lq\cdots\rangle$	$(p\lvert h\rvert l)$	$\sum_q^N \big[(pl\lvert qq) - (pq\lvert ql)\big]$
Case 3	$\lvert\Psi\rangle = \lvert\cdots pq\cdots\rangle$ $\lvert\Phi\rangle = \lvert\cdots lm\cdots\rangle$	0	$\big[(pl\lvert qm) - (pm\lvert ql)\big]$

$$(p|h|l) = \int d\vec{x}_1 \chi_p^*(\vec{x}_1) \left[-\frac{1}{2}\nabla^2 - \sum_A^M \frac{Z_A}{|\vec{r}_1 - \vec{R}_A|}\right] \chi_l(\vec{x}_1)$$

$$(pl|qm) = \int\int d\vec{x}_1 d\vec{x}_2 \chi_p^*(\vec{x}_1)\chi_l(\vec{x}_1) \frac{1}{|\vec{r}_1 - \vec{r}_2|} \chi_q^*(\vec{x}_2)\chi_m(\vec{x}_2)$$

In the table the three cases reported represent respectively the matrix element between SD's that are identical, differ by only one spin orbital (p, l) and differ by two spin orbitals (pq, lm). For two-electron operators, SD's differing by three or more spin orbitals give zero matrix elements. The operators reported here are those that define the Hamiltonian of Equation (2.3), specifically O_1 collects the mono-electronic part of H, that is the sum of the kinetic energy and the electrons-nuclei attraction terms, O_2 is the two-electron part, namely the electron-electron repulsion.

- There is no specification of which electron is in which orbital (the N electrons are indistinguishable).
- Slater determinants formed from orthonormal spin orbitals are normalized.
- Slater determinants made of different orthonormal spin orbitals are orthogonal.
- Interchanging two rows of the SD is equivalent to interchange the coordinates of two electrons. Due to the properties of the determinant this implies the change of sign of the SD, meaning that $\Psi(\vec{x}_1, \ldots, \vec{x}_N)$ expressed in this form fulfill the requirement of the antisymmetry principle.
- Having two equal columns of the SD is equivalent to have two electrons occupying the same spin orbital. Due to the properties of the determinant this implies that the SD is zero, fulfilling the Pauli exclusion principle (no more than one electrons can occupy a spin orbital).

- In the case of N non-interacting electrons (independent electrons) the exact solution can be expressed in the form of a single SD.
- In the case of N interacting electrons the exact solution can be expressed as a linear combination of all possible N-electron Slater determinants formed from a complete set of spin orbitals.

The Hamiltonian given in Equation (2.3) contains only one-electron and two-electron operators. The former is related to the electronic kinetic energy and the Coulomb attraction between electrons and nuclei, the latter is given by the electron-electron repulsion term. In Table 2.1 the rules for calculating matrix elements between Slater Determinants and this kind of operators are summarized.

2.2 Wavefunction Theory

2.2.1 Hartree-Fock Approximation

The Hartree-Fock (HF) approximation represents the basis of most advanced wavefunction theory approaches. In fact the HF method provides the optimized WF that is then used as input for higher level computations and can be considered a preliminary step of these methods.

Within the BO approximation, the Hartree-Fock approach consists in the solution of a multielectronic problem employing as a trial wavefunction a single Slater determinant. Given N electrons, the set of spin orbitals that build up the SD

$$\Psi_0^{\mathrm{HF}} = |\chi_1, \chi_2, \cdots \chi_N\rangle \tag{2.20}$$

are determined variationally by minimizing the expectation value of the Hamiltonian operator

$$E_0^{\mathrm{HF}} = \langle \Psi_0^{\mathrm{HF}} | H | \Psi_0^{\mathrm{HF}} \rangle \tag{2.21}$$

for which the variational freedom is given by the shape of the spin orbitals.

The HF wavefunction Ψ_0^{HF} is thus obtained by minimizing the energy E_0^{HF} with respect to the $\{\chi_p\}$ under the constrain that the spin orbitals remain orthonormal $\langle \chi_p | \chi_q \rangle = \delta_{pq}$. Therefor the optimal spin orbitals are given as the solutions of the Hartree-Fock equations

$$f(\vec{x}_i)\chi(\vec{x}_i) = \epsilon_i \chi(\vec{x}_i) \tag{2.22}$$

$$f(\vec{x}_i) = -\frac{1}{2}\nabla_i^2 - \sum_A \frac{Z_A}{|\vec{r}_i - \vec{R}_A|} + v^{\mathrm{HF}}(\vec{x}_i) \tag{2.23}$$

with $v^{\mathrm{HF}}(\vec{x}_i)$ being the HF mono-electronic potential. These equations are in the form of a mono-electronic eigenvalue problem.

In general, given K spatial orbitals it is possible to generate $2K$ spin orbitals, the ground state wavefunction Ψ_0^{HF} is build with the N lowest solutions of HF equations (occupied molecular orbitals) while the others $2K-N$ remain unoccupied (virtual molecular orbitals). The former are usually labeled with $i, j, k, \ldots$ the later with $a, b, c, \ldots$ indices, while general orbitals are denoted with $p, q, r, \ldots$. The HF limit is reached if the basis of spatial orbitals is complete.

The HF potential $v^{\mathrm{HF}}(\vec{x}_1)$ is made of two terms, namely the Coulomb ($J(\vec{x}_1)$) and Exchange ($K(\vec{x}_1)$) operators, that depends on all occupied spin orbitals solution of the HF equations. They are defined according to

$$v^{\mathrm{HF}}(\vec{x}_1) = \sum_j^N \left[J_j(\vec{x}_1) - K_j(\vec{x}_1)\right] \tag{2.24}$$

$$J_j(\vec{x}_1)\chi_i(\vec{x}_1) = \left[\int d\vec{x}_2 \frac{\chi_j^*(\vec{x}_2)\chi_j(\vec{x}_2)}{|\vec{r}_1 - \vec{r}_2|}\right]\chi_i(\vec{x}_1) \tag{2.25}$$

$$K_j(\vec{x}_1)\chi_i(\vec{x}_1) = \left[\int d\vec{x}_2 \frac{\chi_j^*(\vec{x}_2)\chi_i(\vec{x}_2)}{|\vec{r}_1 - \vec{r}_2|}\right]\chi_j(\vec{x}_1) \tag{2.26}$$

the mono-electronic operator $v^{\mathrm{HF}}(\vec{x}_1)$ can be considered as the effective average potential experienced by an electron in $\vec{x}_1$ due to the presence of all other electrons. With these quantities defined the expectation value of the Hamiltonian, giving the Hartree-Fock energy E_0^{HF}, reads

$$E_0^{\mathrm{HF}} = \langle\Psi_0^{\mathrm{HF}}|H|\Psi_0^{\mathrm{HF}}\rangle = \sum_i^N h_i + \frac{1}{2}\sum_i^N\sum_j^N \left[J_{ij} - K_{ij}\right] \tag{2.27}$$

$$h_i = (i|h|i) = \int d\vec{x}_1 \chi_i^*(\vec{x}_1)\left[-\frac{1}{2}\nabla^2 - \sum_A \frac{Z_A}{|\vec{r}_1 - \vec{R}_A|}\right]\chi_i(\vec{x}_1) \tag{2.28}$$

$$J_{ij} = (ii|jj) = \int\int d\vec{x}_1 d\vec{x}_2 \chi_i^*(\vec{x}_1)\chi_i(\vec{x}_1)\frac{1}{|\vec{r}_1 - \vec{r}_2|}\chi_j^*(\vec{x}_2)\chi_j(\vec{x}_2) \tag{2.29}$$

$$K_{ij} = (ij|ij) = \int\int d\vec{x}_1 d\vec{x}_2 \chi_i^*(\vec{x}_1)\chi_j(\vec{x}_1)\frac{1}{|\vec{r}_1 - \vec{r}_2|}\chi_i^*(\vec{x}_2)\chi_j(\vec{x}_2) \tag{2.30}$$

for which the rules for calculating matrix elements between SD's given in Table 2.1 have been used.

2.2.2 The Roothaan Equations

In the previous section, the HF equations have been presented in the general spin orbital basis. In order to derive a scheme that can be used practically, since in the actual case the Hamiltonian doesn't depend on the spins of the electrons, it is more convenient to work solely on the spatial part of the spin orbitals. This is achieved by integrating out the spin coordinate ω leading to an equivalent set of equations that allows for optimizing only the spatial part of the orbitals.

The derivation is considered only for a closed-shell system, for which there is an even number of electron, half of them with spin up and down respectively. Moreover the spatial part of the spin orbitals is restricted to be equal for the α and β spins

$$\chi_p(\vec{x}) = \psi_p(\vec{r})\alpha(\omega) \tag{2.31}$$

$$\bar{\chi}_p(\vec{x}) = \psi_p(\vec{r})\beta(\omega) \tag{2.32}$$

$$\Psi_0^{\mathrm{RHF}} = |\chi_1, \bar{\chi}_1, \cdots \chi_{\frac{N}{2}}\bar{\chi}_{\frac{N}{2}}\rangle \tag{2.33}$$

where RHF stands for restricted Hartree-Fock. Under these assumptions and thanks to the orthonormalty property of the spin wavefunction, the HF energy is expressed in term of the spatial orbitals as

$$E_0^{\mathrm{RHF}} = 2\sum_i^{N/2} h_i + \frac{1}{2}\sum_i^{N/2}\sum_j^{N/2}[2J_{ij} - K_{ij}] \tag{2.34}$$

$$h_i = \int d\vec{r}_1 \psi_i^*(\vec{r}_1)\left[-\frac{1}{2}\nabla^2 - \sum_A \frac{Z_A}{|\vec{r}_1 - \vec{R}_A|}\right]\psi_i(\vec{r}_1) \tag{2.35}$$

$$J_{ij} = \int\int d\vec{r}_1 d\vec{r}_2 \psi_i^*(\vec{r}_1)\psi_i(\vec{r}_1)\frac{1}{|\vec{r}_1 - \vec{r}_2|}\psi_j^*(\vec{r}_2)\chi_j(\vec{r}_2) \tag{2.36}$$

$$K_{ij} = \int\int d\vec{r}_1 d\vec{r}_2 \psi_i^*(\vec{r}_1)\psi_j(\vec{r}_1)\frac{1}{|\vec{r}_1 - \vec{r}_2|}\psi_i^*(\vec{r}_2)\chi_j(\vec{r}_2) \tag{2.37}$$

and similarly for the Hartree-Fock equations

$$f(\vec{r}_1)\psi_p(\vec{r}_1) = \epsilon_p\psi_p(\vec{r}_1) \tag{2.38}$$

$$f(\vec{r}_1) = h(\vec{r}_1) + \sum_j^{N/2}[2J_j(\vec{r}_1) - K_j(\vec{r}_1)] \tag{2.39}$$

The solution of the Hartree-Fock equations presented here can be accomplished by introducing a basis of K known spatial function $\{\phi_\mu \,|\, \mu = 1, 2, \ldots, K\}$, for example a set of atom-centered Gaussian functions, used to expand the unknown molecular orbitals $\{\psi_p\}$

$$\psi_p(\vec{r}) = \sum_\mu^K \phi_\mu(\vec{r})C_{\mu p} \qquad p = 1, 2, \ldots, K \tag{2.40}$$

this allows to convert the differential equations of 2.38 into a set of algebraic equations that can be solved with standard matrix computations. In fact, inserting Equation (2.40) into Equation (2.38), left-multiplying by $\phi_\nu^*(\vec{r})$ and carrying out the integration leads to

$$\sum_\mu^K C_{\mu p}\int d\vec{r}_1\phi_\nu^*(\vec{r}_1)f(\vec{r}_1)\phi_\mu(\vec{r}_1) = \epsilon_p\sum_\mu^K C_{\mu p}\int d\vec{r}_1\phi_\nu^*(\vec{r}_1)\phi_\mu(\vec{r}_1) \tag{2.41}$$

where the integrals on the left and right hand side are respectively the matrix representation of the Fock operator ($F_{\nu\mu}$) and the overlap matrix elements ($S_{\nu\mu}$) in the basis of the known spatial function $\{\phi_\mu\}$. With this redefinition Equation (2.41) becomes

$$\sum_\mu^K F_{\nu\mu}C_{\mu p} = \epsilon_p\sum_\mu^K S_{\nu\mu}C_{\mu p} \qquad p = 1, 2, \ldots, K \tag{2.42}$$

These are the Roothaan equations [27] and ultimately represent a generalized diagonalization problem, $\mathbf{FC} = \mathbf{SC}\epsilon$, with ϵ being the diagonal matrix of eigenvalues.

The Fock matrix elements in the basis of $\{\phi_\mu\}$ are analogous to that previously defined in the molecular orbital basis, in particular

$$F_{\mu\nu} = h_{\mu\nu} + \sum_{\lambda\sigma}^K P_{\lambda\sigma}\left[(\mu\nu|\lambda\sigma) - \frac{1}{2}(\mu\lambda|\nu\sigma)\right] \tag{2.43}$$

where $h_{\mu\nu}$ is the integral involving the one-electron part of the Fock operator (kinetic energy and nuclear attraction of an electron), $(\mu\nu|\lambda\sigma)$ is a four-index two-electron repulsion integral (ERI) in Mulliken notation

$$(\mu\nu|\lambda\sigma) = \int\int d\vec{r}_1 d\vec{r}_2 \phi_\mu^*(\vec{r}_1)\phi_\nu(\vec{r}_1)\frac{1}{|\vec{r}_1 - \vec{r}_2|}\phi_\lambda^*(\vec{r}_2)\phi_\sigma(\vec{r}_2) \tag{2.44}$$

and $P_{\lambda\sigma}$ are elements of the density matrix depending on the expansion coefficients $C_{\mu i}$ of the occupied orbitals

$$P_{\lambda\sigma} = 2\sum_i^{N/2} C_{\lambda i} C_{\sigma i} \tag{2.45}$$

Therefor, it is clear that the Fock matrix elements depend, through the density matrix, on the expansion coefficients. This implies that the Roothaan equations are nonlinear and thus need to be solved self-consistently, that is, given an initial guess for $\mathbf{P}$, the equations in (2.42) are solved providing a new set of expansion coefficients. These are subsequently used for generating a new Fock matrix and the procedure is iterated till convergence is reached.

Once the Roothaan equations are converged, the Hartree-Fock energy can be computed from $\mathbf{h}, \mathbf{P}$ and $\mathbf{F}$ matrices from the following trace

$$E_0 = \frac{1}{2}\mathrm{Tr}\left[\mathbf{P}(\mathbf{h} + \mathbf{F})\right] \tag{2.46}$$

2.2.3 Excited Determinants and Brillouin's Theorem

For a system with N electrons, given K initial spatial wavefunctions, the (restricted) Hartree-Fock procedure gives $2K$ spin orbitals $\{\chi_p\}$ optimized in a variational sense. The N spin orbitals with lowest eigenvalue ϵ_p are those that build up the best approximation to the ground state wavefunction

$$|\Psi_0\rangle = |\chi_1\chi_2\cdots\chi_i\chi_j\cdots\chi_N\rangle \tag{2.47}$$

in the form a single Slater determinant. This is only one possible way to distribute the N electrons over the $2K$ spin orbitals. The total count of SD's is obtained as the number of all possible combinations of $2K$ elements taken N a time. This is calculated by the binomial coefficient

$$\binom{2K}{N} = \frac{(2K)!}{N!\,(2K-N)!} \tag{2.48}$$

In order to classify these SD's it is convenient to define the HF ground state $|\Psi_0\rangle$ as a reference and label all other determinants according to how they differ from this reference. This labeling is made by specifying which of the N spin orbitals $\{\chi_i \,|\, i = 1, \ldots, N\}$ that build up $|\Psi_0\rangle$, named occupied or hole orbitals, have been replaced by which of the remaining $2K - N$ spin orbitals $\{\chi_a \,|\, a = N+1, \ldots, 2K\}$, *i.e.* the virtual or particle orbitals. These SD's are referred as excited determinants since they represent approximated wavefunctions for the excited state of the considered system.

The excited determinants can thus be classified as singly, doubly, triply, ..., N-tuply excited states respectively if one, two, three, ..., N occupied orbitals of $|\Psi_0\rangle$ have been

replaced by virtual orbitals. Referring to $|\Psi_0\rangle$ given in Equation (2.47), a singly excited determinant for which the occupied spin orbital χ_i is substituted by the virtual χ_a is denoted by

$$|\Psi_i^a\rangle = |\chi_1\chi_2\cdots\chi_a\chi_j\cdots\chi_N\rangle \tag{2.49}$$

In the same manner, a doubly excited determinant for which the occupied χ_i, χ_j are promoted to the virtual χ_a, χ_b is given by

$$|\Psi_{ij}^{ab}\rangle = |\chi_1\chi_2\cdots\chi_a\chi_b\cdots\chi_N\rangle \tag{2.50}$$

and so on for the all higher order excitation.

Although these excited determinants are poor approximations for true excited states, they can be used as an N-electron basis for expanding a multi-determinant wavefunction for improving the ground or excited states description of the system. When doing so, matrix elements of the Hamiltonian over the different Slater determinants $\langle\Psi|H|\Phi\rangle$ need to be computed. Since the Hamiltonian contains only one and two-electron operators, according to the rules given in Table 2.1, determinants that differ more than two spin orbital give zero matrix elements.

In addition to that, Brillouin's Theorem states that singly excited determinants $\{\Psi_i^a\}$ are not directly interacting with the reference Hartree-Fock ground state determinant Ψ_0, that is

$$\langle\Psi_0|H|\Psi_i^a\rangle = 0 \tag{2.51}$$

This can be shown, by first evaluating $\langle\Psi_0|H|\Psi_i^a\rangle$ by using the rules reported in Table 2.1, leading to

$$\langle\Psi_0|H|\Psi_i^a\rangle = (i|h|a) + \sum_j^N \big[(ia|jj) - (ij|ja)\big] \tag{2.52}$$

Analogously, the matrix element $(i|f|a)$ of the mono-electronic Fock operator over the spin orbitals χ_i and χ_a can be calculated by using equations (2.23) and (2.24)

$$(i|f|a) = (i|h|a) + \sum_j^N \big[(ia|jj) - (ij|ja)\big] \tag{2.53}$$

giving the result $\langle\Psi_0|H|\Psi_i^a\rangle = (i|f|a)$. Since both χ_i and χ_a are eigenfunctions of the Fock operator then

$$(i|f|a) = \epsilon_a(i|a) = 0 \tag{2.54}$$

due to the orthonormality of the Hartree-Fock spin orbitals and $i \neq a$, demonstrating the Brillouin's theorem $\langle\Psi_0|H|\Psi_i^a\rangle = (i|f|a) = 0$.

2.2.4 Correlation Energy

The Hartree-Fock model allows to recover roughly $\sim 99\%$ of the total electronic energy. It is capable of correctly accounting for the purely quantum mechanical effect known as Fermi repulsion or exchange correlation, which implies that the probability of finding two electrons with parallel spins at the same point in space is zero. It performs remarkably well in many cases, but it has an important limitation, that is, it completely neglects for

the instantaneous Coulomb repulsion between electrons. In fact, in the HF approximation, each electron feels only the average Coulomb potential arising from all other electrons making the electron-electron repulsion energy too large. Since this term is positive, the HF energy is too high compared to the exact one. This can be understood in simple terms by considering that the electrons will repel each other as they move around, this tendency for the electrons of staying apart is decreasing the repulsion energy.

This simple picture, in which electrons avoid each other, that is, the motion of the electrons in a real system is such that they stay apart rather than being close, is called *electron correlation*. The corresponding energy is formally defined as the difference between the exact non-relativistic electronic energy E_0^{exact} and the Hartee-Fock energy obtained in the limit of a complete basis $E_0^{\text{HF-limit}}$

$$E_{\text{corr}} = E_0^{\text{exact}} - E_0^{\text{HF-limit}} \tag{2.55}$$

Even though the contribution coming from electron correlation constitutes a small fraction of the total energy, accounting for these effects is of prime importance for a quantitative description of chemical and physical phenomena. For this reason several methods have been developed for the accurate treatment of electron correlation. Most of them use the HF wavefunction as starting reference and are often referred to as post-HF methods. Examples of these methods are [28]:

- Many-Body Perturbation Theory (MBPT) based on the Rayleigh-Schrödinger perturbation theory that maps an inexact operator (such as the Fock operator of the HF method), for which the eigenfunctions and eigenvalues are known exactly, to an exact operator, with increasing order of accuracy.
- Configuration Interaction (CI) that uses a trial wavefunction build as a linear combination of Salter determinants (linear ansatz) for which the expansion coefficients are optimized variationally.
- Couple Cluster (CC) for which the trial wavefunction is obtained from an exponential ansatz based on excitation operators.
- Quantum Monte Carlo representing a large class of methods that model the many body effects directly in terms of statistical approaches.

The advantage of these methods is that they create a hierarchical class of approaches for which the level of accuracy can systematically be increased. On the other hand, this comes with the price that also the computational cost grows similarly, ranging from $O(N^4)$ up to exponential, being N a measure of the system size.

2.2.5 Many-Body Perturbation Theory

In perturbation theory the approximated solutions of a given problem are obtained by correcting the exact known solutions of a model problem that differs from the former just by a small perturbation. The application of this approach in quantum mechanics is named Rayleigh-Schrödinger perturbation theory (RSPT) and it is of general validity for any one or N particle system.

As previously stated, assuming to have a zero-order Hamiltonian $H^{(0)}$ for which eigenvalues $E_i^{(0)}$ and eigenfunctions $\Psi_i^{(0)}$ are known exactly

$$H^{(0)}\Psi_i^{(0)} = E_i^{(0)}\Psi_i^{(0)} \tag{2.56}$$

then the problem is to find the solutions of the Hamiltonian H

$$H\Phi_i = \left(H^{(0)} + V\right)\Phi_i = E_i\Phi_i \tag{2.57}$$

where V represents a small perturbation of $H^{(0)}$. The idea of the method is that, if V is somehow small, then Φ_i and E_i are likely close to $\Psi_i^{(0)}$ and $E_i^{(0)}$. Therefor, it is desirable to have a procedure for which the successive corrections to $\Psi_i^{(0)}$ and $E_i^{(0)}$ are systematically going closer and closer to the exact Φ_i and E_i. To do so, a parameter λ is introduced controlling the strength of the perturbation $H' = H^{(0)} + \lambda V$. Since H', through the perturbation parameter, is connecting continuously the zero-order Hamiltonian ($\lambda = 0$) to the true one ($\lambda = 1$), also the relative eigenvalues and eigenfunctions must change continuously. According to this Φ_i and E_i can be written as a Taylor expansion in powers of λ

$$E_i = E_i^{(0)} + \lambda E_i^{(1)} + \lambda^2 E_i^{(2)} + \lambda^3 E_i^{(3)} \dots \tag{2.58}$$

$$\Phi_i = \Psi_i^{(0)} + \lambda\Psi_i^{(1)} + \lambda^2\Psi_i^{(2)} + \lambda^3\Psi_i^{(3)} \dots \tag{2.59}$$

where $E_i^{(n)}$ and $\Psi_i^{(n)}$ represent the nth-order correction to the energy and wavefunction respectively. Eventually λ is set to unity and the nth-order energy or wavefunction is the sum of all contributions up to the nth-order.

Considering normalized unperturbed wavefunctions $\langle\Psi_i^{(0)}|\Psi_i^{(0)}\rangle = 1$, the perturbed wavefunctions Φ_i can be chosen, without loss of generality, to be intermediately normalized, that is $\langle\Phi_i|\Psi_i^{(0)}\rangle = 1$. It follows that

$$\langle\Phi_i|\Psi_i^{(0)}\rangle = \langle\Psi_i^{(0)}|\Psi_i^{(0)}\rangle + \lambda\langle\Psi_i^{(1)}|\Psi_i^{(0)}\rangle + \lambda^2\langle\Psi_i^{(2)}|\Psi_i^{(0)}\rangle + \dots = 1 \tag{2.60}$$

$$\langle\Psi_i^{(n)}|\Psi_i^{(0)}\rangle = 0 \quad \forall n \neq 0 \tag{2.61}$$

where the last equality derives from the fact that Equation (2.60) must be true for any value of λ. The total wavefunction can be subsequently normalized once all corrections up to a given order have been computed.

Substituting Equations (2.58) and (2.59) into $(H^{(0)} + \lambda V)\Phi_i = E_i\Phi_i$ and collecting same terms in λ^n leads to a set of perturbation equations

$$\begin{aligned}
H^{(0)}\Psi_i^{(0)} &= E_i^{(0)}\Psi_i^{(0)} && \lambda^0 \\
H^{(0)}\Psi_i^{(1)} + V\Psi_i^{(0)} &= E_i^{(0)}\Psi_i^{(1)} + E_i^{(1)}\Psi_i^{(0)} && \lambda^1 \\
H^{(0)}\Psi_i^{(2)} + V\Psi_i^{(1)} &= E_i^{(0)}\Psi_i^{(2)} + E_i^{(1)}\Psi_i^{(1)} + E_i^{(2)}\Psi_i^{(0)} && \lambda^2 \\
&\vdots && \vdots \\
H^{(0)}\Psi_i^{(n)} + V\Psi_i^{(n-1)} &= \sum_{k=0}^{n} E_i^{(k)}\Psi_i^{(n-k)} && \lambda^n
\end{aligned} \tag{2.62}$$

Left-multiplying by $\Psi_i^{(0)}$ and carrying out the integration gives the trivial result $E_i^{(0)} = \langle\Psi_i^{(0)}|H^{(0)}|\Psi_i^{(0)}\rangle$ for $n = 0$ while for $n > 0$

$$\begin{aligned}
\langle\Psi_i^{(0)}|H^{(0)}|\Psi_i^{(n)}\rangle + \langle\Psi_i^{(0)}|V|\Psi_i^{(n-1)}\rangle &= \sum_{k=0}^{n} E_i^{(k)}\langle\Psi_i^{(0)}|\Psi_i^{(n-k)}\rangle \\
E_i^{(0)}\langle\Psi_i^{(n)}|\Psi_i^{(0)}\rangle^* + \langle\Psi_i^{(0)}|V|\Psi_i^{(n-1)}\rangle &= \sum_{k=0}^{n} E_i^{(k)}\delta_{kn} \\
\langle\Psi_i^{(0)}|V|\Psi_i^{(n-1)}\rangle &= E_i^{(n)}
\end{aligned} \tag{2.63}$$

for which the relation $\langle\Psi_i^{(0)}|H^0|\Psi_i^{(n)}\rangle = \langle\Psi_i^{(n)}|H^0|\Psi_i^{(0)}\rangle^* = E_i^{(0)}\langle\Psi_i^{(k)}|\Psi_i^{(0)}\rangle^*$ and the intermediately normalization condition, Equation (2.61), have been used. In order for the relations obtained in (2.63) to be practically useful, the unknown $\Psi_i^{(n)}$ have to be expressed in terms of the unperturbed quantities.

The second-order correction to the energy $E_i^{(2)} = \langle\Psi_i^{(0)}|V|\Psi_i^{(1)}\rangle$ is the first term that require for its computation the knowledge of a perturbed quantity, namely $\Psi_i^{(1)}$. In this respect $\Psi_i^{(1)}$ can be obtained by solving the perturbed equation for $n = 1$ in (2.62), that is

$$H^{(0)}\Psi_i^{(1)} + V\Psi_i^{(0)} = E_i^{(0)}\Psi_i^{(1)} + E_i^{(1)}\Psi_i^{(0)} \tag{2.64}$$

Collecting terms in $\Psi_i^{(1)}$ and $\Psi_i^{(0)}$ on the two sides, and substituting $E_i^{(1)} = \langle\Psi_i^{(0)}|V|\Psi_i^{(0)}\rangle$, as obtained from (2.63), leads to

$$(E_i^{(0)} - H^{(0)})\Psi_i^{(1)} = (V - \langle\Psi_i^{(0)}|V|\Psi_i^{(0)}\rangle)\Psi_i^{(0)} \tag{2.65}$$

In order to move on, the first-order correction to the wavefunction $\Psi_i^{(1)}$ is expanded as a linear combination of unperturbed wavefunctions

$$\Psi_i^{(1)} = \sum_k \Psi_k^{(0)} c_{ki}^{(1)} \tag{2.66}$$

where the $\{\Psi_k^{(0)}\}$ is the orthonormal basis formed by the full set of eigenfunctions of $H^{(0)}$. The expansion coefficients $\{c_{ki}^{(1)}\}$ are obtained by projection of $\Psi_i^{(1)}$ on each basis vector $\{\Psi_k^{(0)}\}$

$$c_{ki}^{(1)} = \langle\Psi_k^{(0)}|\Psi_i^{(1)}\rangle \tag{2.67}$$

from which follows, due to the intermediately normalization condition (2.61), that $c_{ii}^{(1)} = 0$. Thus eq. (2.66) can be rewritten as

$$\Psi_i^{(1)} = {\sum_k}' \Psi_k^{(0)}\langle\Psi_k^{(0)}|\Psi_i^{(1)}\rangle \tag{2.68}$$

where the prime is used to highlight that the term $k = i$ is excluded. Substituting this expansion into the equation for the second-order correction to the energy, ($E_i^{(2)} = \langle\Psi_i^{(0)}|V|\Psi_i^{(1)}\rangle$) gives

$$\begin{aligned}
E_i^{(2)} &= \left\langle \Psi_i^{(0)} \middle| V \middle| \left[{\sum_k}' \Psi_k^{(0)}\langle\Psi_k^{(0)}|\Psi_i^{(1)}\rangle \right] \right\rangle \\
&= {\sum_k}' \langle\Psi_i^{(0)}|V|\Psi_k^{(0)}\rangle\langle\Psi_k^{(0)}|\Psi_i^{(1)}\rangle
\end{aligned} \tag{2.69}$$

What remain to be done is to find an expression for $\langle\Psi_k^{(0)}|\Psi_i^{(1)}\rangle$ in terms of unperturbed quantities. This is accomplished by left-multiplying by the $\Psi_k^{(0)}$ eigenfunctions ($k \neq i$) both side of Equation (2.65) and carrying out the integration

$$\begin{aligned}
\langle\Psi_k^{(0)}|E_i^{(0)} - H^{(0)}|\Psi_i^{(1)}\rangle &= \langle\Psi_k^{(0)}|V - \langle\Psi_i^{(0)}|V|\Psi_i^{(0)}\rangle|\Psi_i^{(0)}\rangle \\
E_i^{(0)}\langle\Psi_k^{(0)}|\Psi_i^{(1)}\rangle - \langle\Psi_k^{(0)}|H^{(0)}|\Psi_i^{(1)}\rangle &= \langle\Psi_k^{(0)}|V|\Psi_i^{(0)}\rangle - \langle\Psi_i^{(0)}|V|\Psi_i^{(0)}\rangle\langle\Psi_k^{(0)}|\Psi_i^{(0)}\rangle \\
E_i^{(0)}\langle\Psi_k^{(0)}|\Psi_i^{(1)}\rangle - E_k^{(0)}\langle\Psi_i^{(1)}|\Psi_k^{(0)}\rangle^* &= \langle\Psi_k^{(0)}|V|\Psi_i^{(0)}\rangle - \langle\Psi_i^{(0)}|V|\Psi_i^{(0)}\rangle 0 \\
(E_i^{(0)} - E_k^{(0)})\langle\Psi_k^{(0)}|\Psi_i^{(1)}\rangle &= \langle\Psi_k^{(0)}|V|\Psi_i^{(0)}\rangle \\
\langle\Psi_k^{(0)}|\Psi_i^{(1)}\rangle &= \frac{\langle\Psi_k^{(0)}|V|\Psi_i^{(0)}\rangle}{E_i^{(0)} - E_k^{(0)}}
\end{aligned} \tag{2.70}$$

for which the orthonormality of the zero-order wavefunctions has been used. Substituting the result obtained in (2.70) into Equation (2.69) gives the final working expression for calculating the second-order correction to the energy $E_i^{(2)}$ solely in terms of the perturbation operator V and the known eigenvalues and eigenfunctions of $H^{(0)}$, that is

$$E_i^{(2)} = \sum_k{}' \frac{\langle\Psi_i^{(0)}|V|\Psi_k^{(0)}\rangle\langle\Psi_k^{(0)}|V|\Psi_i^{(0)}\rangle}{E_i^{(0)} - E_k^{(0)}} = \sum_k{}' \frac{\left|\langle\Psi_i^{(0)}|V|\Psi_k^{(0)}\rangle\right|^2}{E_i^{(0)} - E_k^{(0)}} \tag{2.71}$$

The higher order corrections can be obtained employing similar procedures, but with increasing complexity, as an example, the third-order correction for the energy is given by

$$E_i^{(3)} = \sum_{jk}{}' \frac{\langle\Psi_i^{(0)}|V|\Psi_j^{(0)}\rangle\langle\Psi_j^{(0)}|V|\Psi_k^{(0)}\rangle\langle\Psi_k^{(0)}|V|\Psi_i^{(0)}\rangle}{\left(E_i^{(0)} - E_j^{(0)}\right)\left(E_i^{(0)} - E_k^{(0)}\right)} - \langle\Psi_i^{(0)}|V|\Psi_i^{(0)}\rangle \sum_k{}' \frac{\left|\langle\Psi_i^{(0)}|V|\Psi_k^{(0)}\rangle\right|^2}{\left(E_i^{(0)} - E_k^{(0)}\right)^2} \tag{2.72}$$

where again the prime sign is used to express that the terms for which $k = i$ and $j = i$ are excluded from the summations.

The main result here is that, at all orders, the corrections can be obtained from matrix elements of the perturbation operator V over the unperturbed wavefunctions, and the unperturbed energies.

2.2.6 Møller-Plesset Perturbation Theory

The Rayleigh-Schrödinger perturbation theory presented in Section 2.2.5 can be used to improve the Hartree-Fock energy in order to include electron correlation effects, *i.e.* it yields a perturbation expansion for the correlation energy.

In order to do so, the electronic Hamiltonian H is partitioned according to

$$H = H^{(0)} + V \tag{2.73}$$

where $H^{(0)}$ is chosen to be the Hartree-Fock Hamiltonian, obtained as the sum over all N electrons of the mono-electronic Fock operator given in Equation (2.23)

$$H^{(0)} = \sum_i^N f(i) = \sum_i^N [h(i) + v^{\mathrm{HF}}(i)] \tag{2.74}$$

while the perturbation operator V is simply obtained as $V = H - H^{(0)}$ leading to

$$V = \sum_{i<j}^{N} \frac{1}{|\vec{r}_i - \vec{r}_j|} - \sum_{i}^{N} v^{\mathrm{HF}}(i) \tag{2.75}$$

The usage of this partition, together with the general expressions derived from the RSPT, is named Møller-Plesset (MP) Perturbation Theory, and the sequence of the nth-order corrections are denoted with MP1, MP2, MP3, ...

Focusing on the ground state, the zero-order order wavefunction $\Psi_0^{(0)}$ is the Hartree-Fock one with associate zero-order energy $E_0^{(0)}$ given by

$$H^{(0)}\Psi_0^{(0)} = E_0^{(0)}\Psi_0^{(0)} \tag{2.76}$$

$$E_0^{(0)} = \sum_{i}^{N} \epsilon_i \tag{2.77}$$

$$\epsilon_i = (i|f|i) = (i|h|i) + \sum_{j}^{N} \big[(ii|jj) - (ij|ij)\big] \tag{2.78}$$

note that $E_0^{(0)}$ and $\Psi_0^{(0)}$ are respectively the exact eigenvalue and eigenfunction of $H^{(0)}$, additionally $E_0^{(0)}$ is different from the HF energy given in Equation (2.27). The first-order correction to the energy (MP1), according to Equation (2.63), is given by $E_0^{(1)} = \langle\Psi_0^{(0)}|V|\Psi_0^{(0)}\rangle$, that is

$$\begin{aligned}
E_0^{(1)} &= \langle\Psi_0^{(0)}|V|\Psi_0^{(0)}\rangle \\
&= \langle\Psi_0^{(0)}|\sum_{i<j}^{N} \frac{1}{|\vec{r}_i - \vec{r}_j|} - \sum_{i}^{N} v^{\mathrm{HF}}(i)|\Psi_0^{(0)}\rangle \\
&= \langle\Psi_0^{(0)}|\sum_{i<j}^{N} \frac{1}{|\vec{r}_i - \vec{r}_j|}|\Psi_0^{(0)}\rangle - \langle\Psi_0^{(0)}|\sum_{i}^{N} v^{\mathrm{HF}}(i)|\Psi_0^{(0)}\rangle \\
&= \frac{1}{2}\sum_{ij}^{N} \big[(ii|jj) - (ij|ij)\big] - \sum_{i}^{N}(i|v^{\mathrm{HF}}|i) = \\
&= \frac{1}{2}\sum_{ij}^{N} \big[(ii|jj) - (ij|ij)\big] - \sum_{i}^{N}\sum_{j}^{N} \big[(ii|jj) - (ij|ij)\big] \\
&= -\frac{1}{2}\sum_{ij}^{N} \big[(ii|jj) - (ij|ij)\big]
\end{aligned} \tag{2.79}$$

The ground state energy E_0 corrected at the first order is given as the sum of the zero-order energy and the MP1 contribution

$$\begin{aligned}
E_0 = E_0^{(0)} + E_0^{(1)} &= \sum_{i}^{N} \epsilon_i - \frac{1}{2}\sum_{ij}^{N} \big[(ii|jj) - (ij|ij)\big] \\
&= \sum_{i}^{N}(i|h|i) + \frac{1}{2}\sum_{ij}^{N} \big[(ii|jj) - (ij|ij)\big] = E_0^{\mathrm{HF}}
\end{aligned} \tag{2.80}$$

where Equation (2.78) has been used to express the ϵ_i. The E_0 obtained in Equation (2.80) is the Hartree-Fock energy, as given in Equation (2.27), implying that at the first order the MP theory is not yet accounting for the electron correlation, but just correcting for the double counting of the electron-electron interactions in the purely independent particle picture of $H^{(0)}$.

Correlation energy is introduced at the second-order, that is with the $E_0^{(2)}$ contribution (MP2 energy correction). The $E_0^{(2)}$ is given by Equation (2.71), that for the ground state reads

$$E_0^{(2)} = \sum_{k=1} \frac{\left|\langle \Psi_0^{(0)} | V | \Psi_k^{(0)} \rangle\right|^2}{E_0^{(0)} - E_k^{(0)}} \tag{2.81}$$

for which the prime sign has been dropped since by staring from $k = 1$ the ground state is already excluded from the summation. In principle the summation in Equation (2.81) run over all excited Slater determinants that can be obtained from the reference $\Psi_0^{(0)}$. This number, given by Equation (2.48), can be very large even for relatively small systems, but by exploiting the properties of the matrix elements over SD's the summation can be restricted to only the non-zero elements.

First it is noted that the perturbation operator V contains only one and two electron operators, implying that all determinants with excitation order higher than double don't interact with $\Psi_0^{(0)}$, *e.i.* give $\langle \Psi_0^{(0)} | V | \Psi_k^{(0)} \rangle = 0$. Additionally also the singly excited determinants don't interact with $\Psi_0^{(0)}$, in fact

$$\begin{aligned}
\langle \Psi_0^{(0)} | V | \Psi_i^{a(0)} \rangle &= \langle \Psi_0^{(0)} | H - H^{(0)} | \Psi_i^{a(0)} \rangle \\
&= \langle \Psi_0^{(0)} | H | \Psi_i^{a(0)} \rangle - \langle \Psi_0^{(0)} | \sum_j^N f(j) | \Psi_i^{a(0)} \rangle \\
&= 0 - (i|f|a) = 0
\end{aligned} \tag{2.82}$$

where the first term is zero because of the Brillouin's Theorem (Equation (2.51)), the second because χ_i and χ_a are both orthonormal eigenfunctions of the Fock operator (see Equation (2.54)).

In this way the summation in Equation (2.81) is restricted exclusively to the doubly excited states

$$E_0^{(2)} = \sum_{\substack{i<j \\ a<b}} \frac{\left|\langle \Psi_0^{(0)} | V | \Psi_{ij}^{ab(0)} \rangle\right|^2}{E_0^{(0)} - E_{ij}^{ab(0)}} \tag{2.83}$$

where the indices notation $i < j$ and $a < b$ is used to express that the summations run only over unique pairs of occupied and virtual spin orbitals respectively, avoiding double counting of SD's.

The evaluation of the matrix elements at the numerator is accomplished by using the rules given in Table 2.1

$$\begin{aligned}
\langle \Psi_0^{(0)} | V | \Psi_{ij}^{ab(0)} \rangle &= \langle \Psi_0^{(0)} | \sum_{i<j}^N \frac{1}{|\vec{r}_i - \vec{r}_j|} | \Psi_{ij}^{ab(0)} \rangle - \langle \Psi_0^{(0)} | \sum_k^N v^{\mathrm{HF}}(k) | \Psi_{ij}^{ab(0)} \rangle \\
&= \big[(ia|jb) - (ib|ja)\big] - 0
\end{aligned} \tag{2.84}$$

where the second term is zero because v^{HF} is a mono-electronic operator. The unperturbed energies at the denominator are simply given as the sum of all eigenvalues of the Fock operator associated with the spin orbitals that build up the considered SD. With respect to the difference $(E_0^{(0)} - E_{ij}^{ab(0)})$, the only terms that don't cancel each other are the eigenvalues associated with the occupied spin orbitals that are not simultaneously present in $\Psi_0^{(0)}$ and $\Psi_{ij}^{ab(0)}$, that is

$$E_0^{(0)} - E_{ij}^{ab(0)} = \epsilon_i + \epsilon_j - \epsilon_a - \epsilon_b = \Delta_{ij}^{ab} \tag{2.85}$$

Substituting Equations (2.84) and (2.85) into Equation (2.83) leads to

$$E_0^{(2)} = \sum_{\substack{i<j \\ a<b}} \frac{|(ia|jb) - (ib|ja)|^2}{\epsilon_i + \epsilon_j - \epsilon_a - \epsilon_b} \tag{2.86}$$

that represents the second-order correction the the energy expressed in terms of ERI's over spin orbitals and associated eigenvalues. The ground state energy corrected up to the second-order is thus

$$E_0 = E_0^{(0)} + E_0^{(1)} + E_0^{(2)} = E_0^{\mathrm{HF}} + E_0^{(2)} \tag{2.87}$$

The MP2 energy given by Equation (2.86) can be expressed in several more convenient forms. In particular, the constraints $i < j$ and $a < b$ can be dropped by including an extra factor $1/4$ since each term in the summation is symmetric with respect to the exchange of indices $i \leftrightarrow j$ and $a \leftrightarrow b$, and it is zero if $i = j$ or $a = b$

$$E_0^{(2)} = \frac{1}{4} \sum_{ij,ab} \frac{|(ia|jb) - (ib|ja)|^2}{\epsilon_i + \epsilon_j - \epsilon_a - \epsilon_b} \tag{2.88}$$

Considering the short hand notation for the energy denominator Δ_{ij}^{ab}, Equation (2.88) can be further worked out by carrying out the square and collecting identical terms

$$E_0^{(2)} = \frac{1}{4} \sum_{ij,ab} \frac{(ia|jb)^2 - 2(ia|jb)(ib|ja) + (ib|ja)^2}{\Delta_{ij}^{ab}} \tag{2.89}$$

$$= \frac{1}{4} \left[\sum_{ij,ab} \frac{(ia|jb)^2}{\Delta_{ij}^{ab}} - 2 \sum_{ij,ab} \frac{(ia|jb)(ib|ja)}{\Delta_{ij}^{ab}} + \sum_{ij,ab} \frac{(ib|ja)^2}{\Delta_{ij}^{ab}} \right] \tag{2.90}$$

$$= \frac{1}{4} \left[2 \sum_{ij,ab} \frac{(ia|jb)^2}{\Delta_{ij}^{ab}} - 2 \sum_{ij,ab} \frac{(ia|jb)(ib|ja)}{\Delta_{ij}^{ab}} \right] \tag{2.91}$$

$$= \frac{1}{2} \sum_{ij,ab} \frac{(ia|jb)\big[(ia|jb) - (ib|ja)\big]}{\Delta_{ij}^{ab}} \tag{2.92}$$

in Equation (2.90), since the indices run over all allowed states, the first and last summations give equal results.

The formulation for the $E_0^{(2)}$ given in Equation (2.92) is particularly convenient for carrying out the integration over the spin coordinate in order to obtain an expression

depending only on the spatial orbitals. In fact it is clear that the only non vanishing terms in the summation are those for which the integral $(ia|jb)$ is different from zero. Due to the orthonormality of the spin wavefunctions this occurs only in the cases for which the spin orbitals pairs $\chi_i\chi_a$ and $\chi_j\chi_b$ are integrated with the same spin state. In this respect the indices are redefined such that they refer only to the spatial part and, given a generic spatial orbital, the spin is denoted with ψ_p and $\psi_{\bar{p}}$ respectively for the α and β states.

According to this new notation the only non-zero integrals are $(ia|jb)$, $(\overline{ia}|\overline{jb})$, $(\overline{ia}|jb)$ and $(ia|\overline{jb})$. The summation in Equation (2.92) can be spit into these four cases, note that now each index runs only over the spatial functions respectively associated to the α or β spins

$$\begin{aligned} E_0^{(2)} = \frac{1}{2}\Bigg\{ &\sum_{ij,ab} \frac{(ia|jb)\big[(ia|jb)-(ib|ja)\big]}{\Delta_{ij}^{ab}} + \sum_{\overline{ij},\overline{ab}} \frac{(\overline{ia}|\overline{jb})\big[(\overline{ia}|\overline{jb})-(\overline{ib}|\overline{ja})\big]}{\Delta_{\overline{ij}}^{\overline{ab}}} \\ &\sum_{\bar{i}j,\bar{a}b} \frac{(\overline{ia}|jb)\big[(\overline{ia}|jb)-(\bar{i}b|j\bar{a})\big]}{\Delta_{\bar{i}j}^{\bar{a}b}} + \sum_{i\bar{j},a\bar{b}} \frac{(ia|\overline{jb})\big[(ia|\overline{jb})-(i\bar{b}|\bar{j}a)\big]}{\Delta_{i\bar{j}}^{a\bar{b}}} \Bigg\} \end{aligned} \tag{2.93}$$

$$\begin{aligned} = \frac{1}{2}\Bigg\{ &\sum_{ij,ab} \frac{(ia|jb)\big[(ia|jb)-(ib|ja)\big]}{\Delta_{ij}^{ab}} + \sum_{\overline{ij},\overline{ab}} \frac{(\overline{ia}|\overline{jb})\big[(\overline{ia}|\overline{jb})-(\overline{ib}|\overline{ja})\big]}{\Delta_{\overline{ij}}^{\overline{ab}}} \\ &\sum_{\bar{i}j,\bar{a}b} \frac{(\overline{ia}|jb)^2}{\Delta_{\bar{i}j}^{\bar{a}b}} + \sum_{i\bar{j},a\bar{b}} \frac{(ia|\overline{jb})^2}{\Delta_{i\bar{j}}^{a\bar{b}}} \Bigg\} \end{aligned} \tag{2.94}$$

where in the last step the spin coordinates have been integrated out. The last two terms in Equation (2.94) give same result due to the integral symmetry $(\overline{ia}|jb) = (jb|\overline{ia})$ together to the fact that each index runs over all allowed states. This leads to a general expression for $E_0^{(2)}$ that can be used to evaluate the MP2 energy solely in terms of ERI's over the spatial part of spin orbitals and relative eigenvalues

$$E_0^{(2)} = \frac{1}{2}\left\{ \sum_{ij,ab} \frac{(ia|jb)\big[(ia|jb)-(ib|ja)\big]}{\Delta_{ij}^{ab}} + \sum_{\overline{ij},\overline{ab}} \frac{(\overline{ia}|\overline{jb})\big[(\overline{ia}|\overline{jb})-(\overline{ib}|\overline{ja})\big]}{\Delta_{\overline{ij}}^{\overline{ab}}} \right\} + \sum_{\bar{i}j,\bar{a}b} \frac{(\overline{ia}|jb)^2}{\Delta_{\bar{i}j}^{\bar{a}b}} \tag{2.95}$$

note that this equation is valid also in the case for which the α and β spin orbitals have different spatial parts. In Equation (2.95) the terms in curly braces represent the Same-Spin (SS) component while the third is named Opposite-Spin (OS) component of the MP2 energy.

When considering a closed-shell restricted Hartree-Fock reference wavefunction, for which the α and β electrons are paired into orbitals with the same spatial part, that is $\psi_p = \psi_{\bar{p}}$, the first two terms of the SS component are equal. Thus, collecting like integrals, the expression for the MP2 energy reads

$$E_0^{(2)} = \sum_{ij,ab} \frac{(ia|jb)\big[2(ia|jb)-(ib|ja)\big]}{\epsilon_i + \epsilon_j - \epsilon_a - \epsilon_b} \tag{2.96}$$

Equations (2.95) and (2.96) represent the formulas used for the practical implementation of the MP2 energy method. The number of term summed in Equation (2.96) is $o^2 \cdot v^2$,

with o and v being the number of occupied and virtual orbital respectively. This implies that straightforward evaluation of $E_0^{(2)}$ is an $O(N^4)$ scaling method, with N representative of the system size. Despite to that, the real scaling of the MP2 method is $O(N^5)$, due to the atomic orbitals to molecular orbitals integral transformation step, $(\mu\nu|\lambda\sigma) \rightarrow (ia|jb)$. The practical details for the implementation of the MP2 method and related variants will be discussed in the next chapters.

There are many reasons why the MP2 method has been widely and successfully applied in computational chemistry. First of all it represents one of the cheapest post-HF correlation methods, presenting an easy and compact form that generally recovers a large part of the electron correlation energy ($\sim 95\%$). The MP2 energy is size consistent meaning that it leads to additive energies for infinitely separated systems [29]. Most notably, at the MP2 level non-covalent interactions such as van der Waals (dispersion) forces are accounted for. These are purely related to non-local dynamical electron correlation and represent important effects that have to be included for the correct description of many chemical and physical phenomena. For a recent review on this topic see Ref. [30]. The MP2 method is not free of flaws, in particular this approach fails in the cases for which the energy difference between the highest occupied (HOMO) and the lowest unoccupied (LUMO) molecular orbitals approaches zero, making this theory not adequate *e.g.* for the description of metals.

2.2.7 The Random Phase Approximation Correlation Energy

In quantum chemistry the Random Phase Approximation (RPA) has become synonymous with Time Dependent Hartree Fock (TDHF) theory [31, 32]. The latter is an extension of the HF approximation attempting to solve the time-dependent Schrödinger equation in order to describe the absorption of radiation of a given N-electron system. The application of TDHF for calculating correlation energies (including exchange contributions) has been originally worked out by McLachlan and Ball [33]. In this work, a parallelism between TDHF and RPA is shown, the latter was previously developed for describing the correlation energy of the electron gas.

RPA as an approach for calculating correlation energies can be formulated in different way. Following the work of Scuseria and coworkers [34], TDHF as a method for calculating excitation spectra, requires the solution of the following non-Hermitian eigenvalue problem

$$\begin{pmatrix} \mathbf{A} & \mathbf{B} \\ -\mathbf{B} & -\mathbf{A} \end{pmatrix} \begin{pmatrix} \mathbf{X} & \mathbf{Y} \\ \mathbf{Y} & \mathbf{X} \end{pmatrix} = \begin{pmatrix} \mathbf{X} & \mathbf{Y} \\ \mathbf{Y} & \mathbf{X} \end{pmatrix} \begin{pmatrix} \omega & \mathbf{0} \\ \mathbf{0} & -\omega \end{pmatrix} \tag{2.97}$$

where each of the matrices have size $ov \times ov$, being o and v the number of occupied and virtual spin orbitals. The positive and negative eigenvalues ω_i represent respectively excitation and de-excitation energies with corresponding eigenvectors $\binom{\mathbf{X}_i}{\mathbf{Y}_i}$ and $\binom{\mathbf{Y}_i}{\mathbf{X}_i}$. The matrices $\mathbf{A}$ and $\mathbf{B}$ are defined in the spin orbital basis as

$$A_{ia,jb} = (\epsilon_a - \epsilon_i)\delta_{ij}\delta_{ab} + (ia|jb) - (ij|ab) \tag{2.98}$$

$$B_{ia,jb} = (ia|jb) - (ib|ja) \tag{2.99}$$

with the usual convention for the molecular orbital energies, ERI's and occupied and virtual indices.

The RPA correlation energy can be obtained by considering two (harmonic) excitation energy problems: TDHF and Configuration Interaction Singles (CIS). The latter solves the problem in Equation (2.97) by setting $\mathbf{B} = 0$

$$\mathbf{AZ} = \mathbf{Z}\nu \tag{2.100}$$

CIS, due to the Brillouin's Theorem, doesn't correlate the ground state. Contrary to that, TDHF gives correlation of the ground state. In this respect the RPA correlation energy is given as the difference between the zero point energies of these two harmonic oscillator problems with correlated (TDHF) and uncorrelated (CIS) ground states, that is

$$E_c^{\mathrm{RPA}} = \frac{1}{2}\sum_i(\omega_i - \nu_i) = \frac{1}{2}\mathrm{Tr}(\omega - \mathbf{A}) \tag{2.101}$$

this equation is also known as Plasmon equation [35].

The RPA correlation energy formulated here is referred as RPA *exchange* (RPAx). More common is the *direct* RPA (dRPA) method that is obtained in the same way as RPAx but neglecting all exchange integrals in the $\mathbf{A}$ and $\mathbf{B}$ matrices, that is only the $(ia|jb)$ are retained, but not $(ib|ja)$ and $(ij|ab)$.

The RPAx based on Hartree-Fock has been rarely used for calculating the correlation energy, partially because the higher computational cost compared for example to MP2 and partially because its poor performance due to the triplet instabilities of the HF reference [36]. The dRPA correlation energy approach, for which low order scaling implementations have been developed, is in general superior than RPAx and much more common. Additionally dRPA has a closer connection to the Kohn-Sham method in the framework of Density Functional Theory.

2.3 Density Functional Theory

Density Functional Theory (DFT), after the introduction of the Kohn-Sham (KS) method, has become the most widely used quantum mechanical approach for studying a large variety of problems ranging from molecular to condensed matter applications. The reason is in the excellent compromise between computational cost and performance, very competitive compared to costly wavefunction based methods. DFT has also been extensively used in this thesis work and for this reason the main features of this approach are briefly recalled in this section. Additionally the RPA correlation energy method is generally based on KS-DFT references, for which a close connection is established in terms of the adiabatic-connection fluctuation-dissipation theorem.

The basic premise of DFT is that all the intricate motions and pair correlations in a many-electron system are somehow completely contained in the total electronic density. In this respect it appears clear that changing the point of view from the complicated N-body wavenfunction $\Psi(\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_N)$ to the much simple picture of electronic density $\rho(\vec{r})$ is very appealing [37]. Central in DFT is the concept of functional that represents an application that associates a scalar value to a multidimensional function. In DFT the total electronic energy of the system is given as a functional of the total electronic density.

In the non-relativistic case, within the Born-Oppenheimer approximation, the Hamiltonian of an N-electron system can be formulated as

$$H = -\frac{1}{2}\sum_{i=1}^{N}\nabla_i^2 + \sum_{i<j}^{N}\frac{1}{r_{ij}} + \sum_{i=1}^{N}V_{ext}(\vec{r}_i) \tag{2.102}$$

where the terms on the right hand side represent respectively the electron kinetic energy, the electron-electron repulsion energy (with r_{ij} shorthand notation for $|\vec{r}_i - \vec{r}_j|$) and the Coulomb attraction between the electrons and the external potential (usually due to the nuclei). The Hamiltonian, therefore, can be expressed as the sum of three operators, each of them associated to the energy contributions introduced in Equation (2.102):

$$H = T + V_{ee} + V_{ext} \tag{2.103}$$

from which it appears that H is completely fixed once V_{ext} is defined.

Given an arbitrary antisymmetric N-electron wavefunction $\Psi(\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_N)$, the expectation value of the external potential V_{ext} can be written as the following integral

$$\langle\Psi|\, V_{ext}\,|\Psi\rangle = \int_V \rho(\vec{r})\cdot V_{ext}(\vec{r})d\vec{r} \tag{2.104}$$

where the total electronic density $\rho(\vec{r})$ is introduced and defined as:

$$\rho(\vec{r}) = N\int \Psi^*(\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_N)\Psi(\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_N)d\omega_1 d\vec{x}_2 \ldots d\vec{x}_N \tag{2.105}$$

In Equation (2.105) the first variable is integrated only over the spin coordinate ω, giving a function that delivers the total electronic density at each point in space $\vec{r}$. If the exact ground state wavefunction Ψ_0 is known, then Equation (2.105) provides the exact ground state electronic density ρ_0.

According to this basic definition, it is possible to introduce an universal functional $F[\rho]$ as:

$$F[\rho] = \underset{\Psi\to\rho}{min}\,\langle\Psi|\, T + V_{ee}\,|\Psi\rangle \tag{2.106}$$

where $\underset{\Psi\to\rho}{min}$ means that, given an arbitrary electronic density ρ, the wavefunction Ψ is selected among all $\{\Psi\}$ that for quadrature give ρ (Equation (2.105)), such that the associated expectation value $\langle\Psi|\, T + V_{ee}\,|\Psi\rangle$ is minimum. The quantity $\langle\Psi^{\rho}_{min}|\, T + V_{ee}\,|\Psi^{\rho}_{min}\rangle$ gives the value to the functional. $F[\rho]$ defines an universal functional because it is independent from the external potential V_{ext}. Combining Equation (2.104) and Equation (2.106) allows to define the energy functional $E[\rho]$

$$E[\rho] = F[\rho] + \int_V \rho(\vec{r})\cdot V_{ext}(\vec{r})d\vec{r} \tag{2.107}$$

The mathematical foundation of DFT lies in the two Hohenberg-Kohn theorems [1], that respectively lead to the following conclusions

1. There is a one-to-one correspondence between the exact ground state electronic density ρ_0 and the external potential V_{ext}. Since V_{ext} fixes the Hamiltonian H, then the full many particle ground state is a unique functional of ρ_0.

2. The variational principle can be reformulated in terms of the energy functional $E[\rho]$ defined in Equation (2.107) and the electronic density ρ

$$E[\rho] \geq E_0 \tag{2.108}$$

$$E[\rho_0] = E_0 \tag{2.109}$$

where ρ is an arbitrary electronic density, ρ_0 is the exact ground state density and E_0 is the exact ground state energy.

Since $F[\rho]$ associates to ρ the corresponding Ψ^{ρ}_{min} then

$$F[\rho] = \overset{min}{\Psi \to \rho} \langle \Psi | \, T + V_{ee} \, | \Psi \rangle = \langle \Psi^{\rho}_{min} | \, T + V_{ee} \, | \Psi^{\rho}_{min} \rangle \tag{2.110}$$

from which it is possible to define two other functionals related to the electronic kinetic energy and to the electron-electron repulsion:

$$T[\rho] = \langle \Psi^{\rho}_{min} | \, T \, | \Psi^{\rho}_{min} \rangle \tag{2.111}$$

$$V_{ee}[\rho] = \langle \Psi^{\rho}_{min} | \, V_{ee} \, | \Psi^{\rho}_{min} \rangle \tag{2.112}$$

So the task of minimizing the energy with respect to the electronic density ultimately results into the minimization of the functional

$$E[\rho] = T[\rho] + V_{ee}[\rho] + \int \rho(\vec{r}) \cdot V_{ext}(\vec{r}) d\vec{r} \tag{2.113}$$

with respect to ρ. In this respect two approaches are possible:

- Direct method (Thomas-Fermi approach), where the electronic density is varied.
- Indirect method (Kohn-Sham approach), where orbitals are introduced and varied.

In principle, independently on the minimization approach used, for a given external potential, the final optimized ρ that minimize $E[\rho]$ should deliver the exact ground state electronic density and energy. Unfortunately complications arise from the fact that the explicit forms of the two functionals $T[\rho]$ and $V_{ee}[\rho]$ are not known and approximations have to be considered.

For $V_{ee}[\rho]$ it can be argued that large part of the the electron-electron repulsion energy is associated to the classical Hartree self interaction of $\rho(\vec{r})$ while the remaining (small) unknown contributions can be collected into a functional $E_{\mathrm{NC}}[\rho]$ representative of all non-classical effects

$$V_{ee}[\rho] = J[\rho] + E_{\mathrm{NC}}[\rho] \tag{2.114}$$

$$J[\rho] = \frac{1}{2} \int \int \frac{\rho(\vec{r}) \rho(\vec{r}\,')}{|\vec{r} - \vec{r}\,'|} d\vec{r} d\vec{r}\,' \tag{2.115}$$

A similar simple partition is not possible for $T[\rho]$, additionally accurate models for the kinetic energy functional in term solely of the density are hard to be constructed. The Kohn-Sham approach, by introducing an orbital basis, allows for a simplified and accurate treatment of $T[\rho]$ and that's the reason why the KS method represented a breakthrough in DFT.

2.3.1 Kohn-Sham Approach

The basis of the KS method lies in the introduction of a reference system made of N non-interacting electrons ($V_{ee} = 0$) defined such that its density ρ is the same as that of the fully interacting real system. The Hamiltonian of such a system (the Kohn-Sham Hamiltonian) reads:

$$H_{\mathrm{KS}} = -\frac{1}{2}\sum_{i=1}^{N}\nabla_i^2 + \sum_{i=1}^{N} V_{\mathrm{KS}}(\vec{r}_i) \tag{2.116}$$

where $V_{\mathrm{KS}}(\vec{r}_i)$ is a local mono-electronic potential chosen such that the previously mentioned constraint on the electronic density is fulfilled.

The exact solution of an Hamiltonian associated to a system made up of non-interacting electrons can be chosen in the form of a single Slater Determinant, therefore the solution the Kohn-Sham Hamiltonian can be expressed mathematically as:

$$\Psi_{\mathrm{KS}} = |\varphi_1\varphi_2\ldots\varphi_N\rangle \tag{2.117}$$

with the usual requirement that the spin orbitals are orthonormal $\langle\varphi_i|\varphi_j\rangle = \delta_{ij}$. If $V_{\mathrm{KS}}(\vec{r}_i)$ is known, since H_{KS} is the sum of single particle operators, the optimal spin orbitals are obtained solving:

$$\left[-\frac{1}{2}\nabla^2 + V_{\mathrm{KS}}\right]\varphi_i = \epsilon_i\varphi_i \tag{2.118}$$

analogously to the Hartree-Fock equations. By construction the N lowest solutions of Equation (2.118), after integrating out the spin coordinate, fulfill

$$\rho(\vec{r}) = \sum_{i=1}^{N}|\varphi_i(\vec{r})|^2 \tag{2.119}$$

$\rho(\vec{r})$ being the same electronic density as that of the interacting system. Once the optimized set of spin orbitals $\{\varphi_i\}$ is made available, it can be used to approximate the kinetic energy functional for the interacting system

$$T_0[\rho] = \sum_{i=1}^{N}\langle\varphi_i| - \frac{1}{2}\nabla^2|\varphi_i\rangle \tag{2.120}$$

This represents one of the strength of the Kohn-Sham approach, *i.e.* the way of calculating the kinetic energy, that has proven to be a very good approximation. Therefor the energy functional for the non interacting system reads

$$E[\rho] = T_0[\rho] + \int \rho(\vec{r})\cdot V_{\mathrm{KS}}(\vec{r})d\vec{r} \tag{2.121}$$

For the interacting system, the energy functional is partitioned in different contributions:

$$E[\rho] = T_0[\rho] + J[\rho] + E_{xc}[\rho] + \int \rho(\vec{r})\cdot V_{ext}(\vec{r})d\vec{r} \tag{2.122}$$

where $J[\rho]$ is the classical Hartree self interaction of $\rho(\vec{r})$ as given in Equation (2.115). In Equation (2.122) all unknown quantities are collected into $E_{xc}[\rho]$, namely the *exchange correlation* functional, formally defined as

$$E_{xc}[\rho] = (T[\rho] - T_0[\rho]) + (V_{ee}[\rho] - J[\rho]) \tag{2.123}$$

for which the terms in parenthesis give respectively the difference between the kinetic energy of the interacting and non-interacting system and the deviation from the classical mean field Coulomb interaction due to the quantum nature and pair correlation of the electrons. Even though the explicit form of $E_{xc}[\rho]$ is unknown, this represent a small fraction of the total energy and can thus be approximated using various models referred as density functional approximations (DFA).

The Kohn-Sham equations [2] are obtained by minimizing the energy functional $E[\rho]$ with respect to ρ under the constraint that the number of electron N remains constant, that is

$$\int \rho(\vec{r}) d\vec{r} - N = 0 \tag{2.124}$$

This constrained minimization is accomplished by introducing the multiplier μ and setting up the Lagrangian

$$L = E[\rho] - \mu \left(\int \rho(\vec{r}) d\vec{r} - N \right) \tag{2.125}$$

from which the stationary condition becomes

$$\frac{\delta}{\delta \rho} \left[E[\rho] - \mu \int \rho(\vec{r}) d\vec{r} \right] = 0 \tag{2.126}$$

According to the theory of the functional derivatives, Equation (2.126) leads

- for the non-interacting system (Equation 2.121)

$$\frac{\delta T_0}{\delta \rho} + V_{\mathrm{KS}}(\vec{r}) = \mu \tag{2.127}$$

- for the interacting system (Equation 2.122)

$$\frac{\delta T_0}{\delta \rho} + \int \frac{\rho(\vec{r}\,')}{|\vec{r} - \vec{r}\,'|} d\vec{r}\,' + \frac{\delta E_{xc}}{\delta \rho} + V_{ext}(\vec{r}) = \mu \tag{2.128}$$

Since the same electronic density ρ is imposed for the interacting and non-interacting systems, from Equation (2.127) and Equation (2.128) an expression for the Kohn-Sham potential V_{KS} is found

$$V_{\mathrm{KS}}(\vec{r}) = \int \frac{\rho(\vec{r}\,')}{|\vec{r} - \vec{r}\,'|} d\vec{r}\,' + \frac{\delta E_{xc}}{\delta \rho} + V_{ext}(\vec{r}) \tag{2.129}$$

according to this V_{KS} depends on the classical Coulomb repulsion associated to the electronic density ρ (first term), the functional derivative of E_{xc} (second term) and the external potential (last term). Since a general prescription on how to form V_{KS} is achieved, the optimized spin orbitals that build up the single Slater Determinant that minimize

the functional $E[\rho]$, are obtained by solving the Equation (2.118), that represents the Kohn-Sham equations.

In summary, given an N-electron system under the influence of an external potential $V_{ext}(\vec{r})$, the KS-DFT method allows to optimize a set of orthonormal spin orbitals, obtained as the solution of the KS equations, from which it is possible to calculated the exact ground state electronic density ρ_0 and energy E_0 if the exact form of the exchange correlation potential $V_{xc}(\vec{r})$ is known

$$\begin{gathered}
\{\varphi_i(\vec{x}) \,|\, i = 1, 2, \ldots, N\} \qquad \langle\varphi_i|\varphi_j\rangle = \delta_{ij} \\
h^{\mathrm{KS}}\varphi_i = \left[-\frac{1}{2}\nabla^2 + V_{\mathrm{KS}}\right]\varphi_i = \epsilon_i\varphi_i \\
V_{\mathrm{KS}}(\vec{r}) = V_H(\vec{r}) + V_{xc}(\vec{r}) + V_{ext}(\vec{r}) \\
V_H(\vec{r}) = \int \frac{\rho(\vec{r}\,')}{|\vec{r} - \vec{r}\,'|} d\vec{r}\,' \quad , \quad V_{xc}(\vec{r}) = \frac{\delta E_{xc}[\rho]}{\delta\rho} \\
\rho_0(\vec{r}) = \sum_{i=1}^{N} \int |\varphi_i(\vec{x})|^2 d\omega \\
E_0 = E[\rho_0] = -\frac{1}{2}\sum_{i=1}^{N} \langle\varphi_i| \nabla^2 |\varphi_i\rangle + \frac{1}{2}\int\int \frac{\rho_0(\vec{r})\rho_0(\vec{r}\,')}{|\vec{r} - \vec{r}\,'|} d\vec{r} d\vec{r}\,' \\
+ \int \rho_0(\vec{r}) \cdot V_{xc}(\vec{r}) d\vec{r} + \int \rho_0(\vec{r}) \cdot V_{ext}(\vec{r}) d\vec{r}
\end{gathered} \tag{2.130}$$

Since the exact explicit form of $V_{xc}(\vec{r})$ is unknown, approximations have to be introduced. It is clear that the accuracy of the calculated energy and density depends on the quality of the exchange correlation potential, that is the level of sophistication of the employed model for $V_{xc}(\vec{r})$. Contrary to wavefunction theory, for which a class of approaches allowing for a systematic improvement of the electronic structure description exists, in DFT this is hard to be established due to the lack of a general methodology for the construction of $V_{xc}(\vec{r})$. Despite this Perdew *et al.* [3] classified the density functional approximation on a "Jacob's ladder" for which each rung of the ladder introduces more descriptors of the electronic system and yields models with improved accuracy. At the present stage, this ladder have five rungs making use progressively of the information on 1) the electronic density, 2) its gradient, 3) the kinetic energy density, 4) the occupied molecular orbitals (MO) and 5) the virtual MO's.

2.3.2 Practical Solution of the Kohn-Sham Equations

In most of the applications of the KS-DFT method, the Kohn-Sham equations are solved by introducing a basis of known spatial function used to expand the spatial part of the spin orbital eigenfunctions of the KS Hamiltonian. Similar to the case of the Hartree-Fock equations this allow for reformulating the problem summarized in (2.130) in terms of a set of algebraic equations that can be solved with standard matrix computations. When atom centered functions are used as a basis this approach is also known as linear combination of atomic orbitals (LCAO) expansion.

Given a set of K predefined spatial basis functions $\{\psi_\nu\}$ then the spatial part of the Kohn-Sham orbitals are expand as a linear combination

$$\varphi_i(\vec{r}) = \sum_{\nu=1}^{K} \psi_\nu(\vec{r}) C_{\nu i} \tag{2.131}$$

Inserting Equation (2.131) into the KS equations (second line of 2.130), multiplying from the left with an arbitrary basis function ψ_μ and integrating, leads to a series of K equations

$$\sum_{\nu=1}^{K} C_{\nu i} \int \psi_\mu^*(\vec{r})\, h^{\mathrm{KS}}(\vec{r})\, \psi_\nu(\vec{r}) d\vec{r} = \epsilon_i \sum_{\nu=1}^{K} C_{\nu i} \int \psi_\mu^*(\vec{r})\, \psi_\nu(\vec{r}) d\vec{r} \tag{2.132}$$
$$1 \leq i \leq K$$

for which the integrals on both sides of this equation correspond to matrix elements. On the left hand side

$$H_{\mu\nu}^{\mathrm{KS}} = \int \psi_\mu^*(\vec{r}) \left[-\frac{1}{2}\nabla^2 + V_{\mathrm{KS}}(\vec{r}) \right] \psi_\nu(\vec{r}) d\vec{r} \tag{2.133}$$

is a matrix element of the *Kohn-Sham matrix* while the *overlap matrix* on the right hand side has elements

$$S_{\mu\nu} = \int \psi_\mu^*(\vec{r})\, \psi_\nu(\vec{r}) d\vec{r} \tag{2.134}$$

Both these matrices ($\mathbf{S}$ and $\mathbf{H}^{\mathrm{KS}}$) are square and symmetric with dimension $K \times K$. Introducing the $\mathbf{C}$ matrix containing the expansion vectors and $\mathbf{E}$, as the diagonal matrix of the orbital energies ϵ_i, it is possible to rewrite the K Equation (2.132) compactly as a matrix equation:

$$\mathbf{H}^{\mathrm{KS}}\mathbf{C} = \mathbf{SCE} \tag{2.135}$$

The solution of this matrix equation represent a standard generalized diagonalization problem. Since $\mathbf{H}^{\mathrm{KS}}$ through $V_{\mathrm{KS}}(\vec{r})$ depends on the electronic density and so for $\mathbf{C}$, the solution of this problem requires an iterative procedure named *self consistent field* (SCF).

The Kohn-Sham potential $V_{\mathrm{KS}}(\vec{r})$ contains the Hartree potential term $V_H(\vec{r})$ that requires the integration over two spatial coordinates. Despite that, by using standard techniques to solve the Poisson equation, the electrostatic potential $V_H(\vec{r})$ can be directly calculated from the electronic density avoiding the calculation of integrals over two spatial coordinates. In this way, $V_{\mathrm{KS}}(\vec{r})$ can be expressed in a fully local form and this makes the computational cost for the construction of $\mathbf{H}^{\mathrm{KS}}$, in the worst case, an $O(N^3)$ operation.

Due to the complicated nature of the potentials building up $V_{\mathrm{KS}}(\vec{r})$, these integral are usually evaluated numerically. In this respect, efficient molecular quadrature schemes have been developed [38–40].

The same $O(N^3)$ effort is necessary for solving the matrix equation (2.135), resulting in an $O(N^3)$ final computational cost for the overall KS-DFT procedure. This can be considered as an upper bound, since efficient linear scaling DFT approaches have been developed and implemented, contributing to the wide spread use of KS-DFT method.

2.3.3 Reduced Density Matrices and Exchange-Correlation Hole

The quest of finding approximations for $E_{xc}[\rho]$ can be more conveniently accomplished by introducing the concepts of reduced density matrices rather than using solely the electronic density. Ignoring for practicality the electron spin, the first (γ_1) and second (γ_2) order reduced density matrices are defined as

$$\gamma_1(\vec{r}_1, \vec{r}_1\,') = N \int \Psi^*(\vec{r}_1\,', \vec{r}_2, \ldots, \vec{r}_N)\Psi(\vec{r}_1, \vec{r}_2, \ldots, \vec{r}_N)d\vec{r}_2, \ldots, d\vec{r}_N \tag{2.136}$$

$$\gamma_2(\vec{r}_1, \vec{r}_2, \vec{r}_1\,', \vec{r}_2\,') = N(N-1) \int \Psi^*(\vec{r}_1\,', \vec{r}_2\,', \vec{r}_3, \ldots, \vec{r}_N)\Psi(\vec{r}_1, \vec{r}_2, \vec{r}_3, \ldots, \vec{r}_N)d\vec{r}_3, \ldots, d\vec{r}_N \tag{2.137}$$

obtained by integrating out all electron coordinates except one and two for γ_1 and γ_2 respectively.

The diagonal part of $\gamma_1(\vec{r}_1, \vec{r}_1\,')$, *i.e.* the elements corresponding to $\vec{r}_1 = \vec{r}_1\,'$ represents the electronic density $\rho_1(\vec{r}) = \gamma_1(\vec{r}_1, \vec{r}_1)$ (first-order density matrix) as defined in Equation (2.105), representing the probability of finding an electron at any position $\vec{r}$ in space. According to this, the prefactor N ensures that the total number of electrons is recovered by integration of $\rho_1(\vec{r})$.

Analogously, the diagonal elements of $\gamma_2(\vec{r}_1, \vec{r}_2, \vec{r}_1\,', \vec{r}_2\,')$, that is $\vec{r}_1 = \vec{r}_1\,'$ and $\vec{r}_2 = \vec{r}_2\,'$, give the second-order density matrix or electron pair-density $\rho_2(\vec{r}_1, \vec{r}_2)$. This function is normalized to the number of non-distinct electron pairs $N(N-1)$ and represents the probability of finding two electrons respectively at position $\vec{r}_1$ and $\vec{r}_2$. If Ψ is the exact ground state wavefunction, then $\rho_2(\vec{r}_1, \vec{r}_2)$ contains all information about electron correlation. The electron-electron repulsion energy V_{ee} can in fact be expressed in terms of γ_2 and ρ_2 as

$$V_{ee} = \frac{1}{2} \int \frac{\gamma_2(\vec{r}_1, \vec{r}_2, \vec{r}_1, \vec{r}_2)}{|\vec{r}_1 - \vec{r}_2|} d\vec{r}_1 d\vec{r}_2 = \frac{1}{2} \int \frac{\rho_2(\vec{r}_1, \vec{r}_2)}{|\vec{r}_1 - \vec{r}_2|} d\vec{r}_1 d\vec{r}_2 \tag{2.138}$$

With these quantities defined, also the the exchange-correlation energy can be reformulated in terms of density matrices. For this purpose, the concept of *exchange* and *correlation holes* is introduced, that can be pictorially interpreted by considering that each electron is surrounded by an hole that diminishes the probability of finding another electron in its vicinity.

If the N electrons are completely independent particles, then the probability of finding an electron at any position in space is uncorrelated to the position of any other electron, meaning that the pair-density ρ_2 is simply given as the product of two one-electron densities ρ_1

$$\rho_2^{\text{indep}}(\vec{r}_1, \vec{r}_2) = \frac{N-1}{N} \rho_1(\vec{r}_1)\rho_1(\vec{r}_2) = \left(1 - \frac{1}{N}\right) \rho_1(\vec{r}_1)\rho_1(\vec{r}_2) \tag{2.139}$$

Here the factor $(N-1)/N$ is introduced because, since electrons are indistinguishable, the probability of finding an electron in $\vec{r}_1$ is given by $\rho_1(\vec{r}_1)$, while for a second electron in $\vec{r}_2$ the probability has to be scaled by $(N-1)/N$ since the electron at $\vec{r}_1$ can not be simultaneously also at $\vec{r}_2$.

For the real interacting system, the probability of finding an electron near another is reduced. This can be formulated by introducing a conditional probability factor $h_{xc}(\vec{r}_1, \vec{r}_2)$, including the $1/N$ self-interaction factor in Equation (2.139), giving

$$\rho_2(\vec{r}_1, \vec{r}_2) = \rho_1(\vec{r}_1)\rho_1(\vec{r}_2) + \rho_1(\vec{r}_1)h_{xc}(\vec{r}_1, \vec{r}_2) \tag{2.140}$$

The conditional probability $h_{xc}(\vec{r}_1, \vec{r}_2)$ is called *exchange-correlation hole*, and is formally defined in terms of ρ_1 and ρ_2 by inverting Equation (2.140)

$$h_{xc}(\vec{r}_1, \vec{r}_2) = \frac{\rho_2(\vec{r}_1, \vec{r}_2)}{\rho_1(\vec{r}_1)} - \rho_1(\vec{r}_2) \tag{2.141}$$

The exchange-correlation hole $h_{xc}(\vec{r}_1, \vec{r}_2)$ represents the reduced probability of finding an electron in $\vec{r}_1$ given a second electron in $\vec{r}_2$.

Given the electron repulsion energy as given in Equation (2.138) and the definition of ρ_2 in term of $h_{xc}(\vec{r}_1, \vec{r}_2)$ in Equation (2.140) allows to write V_{ee} as

$$\begin{aligned} V_{ee} &= \frac{1}{2}\int \frac{\rho_1(\vec{r}_1)\rho_1(\vec{r}_2)}{|\vec{r}_1 - \vec{r}_2|} d\vec{r}_1 d\vec{r}_2 + \frac{1}{2}\int \frac{\rho_1(\vec{r}_1)h_{xc}(\vec{r}_1, \vec{r}_2)}{|\vec{r}_1 - \vec{r}_2|} d\vec{r}_1 d\vec{r}_2 \\ &= J[\rho] + \frac{1}{2}\int \frac{\rho_1(\vec{r}_1)h_{xc}(\vec{r}_1, \vec{r}_2)}{|\vec{r}_1 - \vec{r}_2|} d\vec{r}_1 d\vec{r}_2 \end{aligned} \tag{2.142}$$

2.3.4 The Adiabatic Connection

The KS scheme is based on the introduction of two N-electron systems namely the non-interacting reference, for which $V_{ee} = 0$, and the real one with full electron pair interactions. In this respect, it is possible to imagine a path that connects these two systems and thus allows to slowly convert the non-interacting into the real system. In order to do so a *coupling strength parameter* λ is introduced that by varying from 0 to 1 converts the non-interacting to the fully interacting system.

The electronic energy along this path is obtained from the associated λ-dependent Hamiltonian H^λ

$$H^\lambda = T + V_{ext}^\lambda + \lambda \sum_{i<j} \frac{1}{r_{ij}} \tag{2.143}$$

$$E^\lambda = \langle \Psi^\lambda | \, H^\lambda \, | \Psi^\lambda \rangle \tag{2.144}$$

where Ψ^λ is the ground state wavefunction of H^λ and V_{ext}^λ is and effective external potential defined such that for any value of λ the electronic density of the system ρ is the same as that of the fully interacting system. If $\lambda = 0$ then $H^{\lambda=0}$ recovers the KS Hamiltonian with $V_{ext}^{\lambda=0} = V_{\mathrm{KS}}$ being the KS effective potential, while, for $\lambda = 1$, $V_{ext}^{\lambda=1}$ reduces to the bare external potential. Equation (2.143) describes how the system is smoothly connected between the two extreme cases, the associated path is referred as the *adiabatic connection*.

The energy difference between the extreme cases $\lambda = 1$ and $\lambda = 0$ can be expressed in an integral form as

$$E^{\lambda=1} - E^{\lambda=0} = \int_{\text{non-interacting}}^{\text{fully-interacting}} dE^\lambda = \int_0^1 \frac{dE^\lambda}{d\lambda} d\lambda \tag{2.145}$$

from which it follows that

$$E^{\lambda=1} = \int_0^1 \frac{dE^\lambda}{d\lambda} d\lambda + E^{\lambda=0} \tag{2.146}$$

The derivative of E^λ with respect to λ is obtained from Equations (2.143) and (2.144) by applying the Hellmann-Feynman theorem

$$\begin{aligned} \frac{dE^\lambda}{d\lambda} &= \frac{d}{d\lambda} \langle \Psi^\lambda | H^\lambda | \Psi^\lambda \rangle = \langle \Psi^\lambda | \frac{dH^\lambda}{d\lambda} | \Psi^\lambda \rangle \\ &= \langle \Psi^\lambda | \frac{dV_{ext}^\lambda}{d\lambda} + V_{ee} | \Psi^\lambda \rangle \end{aligned} \tag{2.147}$$

Substituting Equation (2.147) into Equation (2.146) gives

$$E^{\lambda=1} = \int_0^1 \langle \Psi^\lambda | \frac{dV_{ext}^\lambda}{d\lambda} | \Psi^\lambda \rangle \, d\lambda + \int_0^1 \langle \Psi^\lambda | V_{ee} | \Psi^\lambda \rangle \, d\lambda + E^{\lambda=0} \tag{2.148}$$

since the electronic density is constrained to be the same over the integration path, that is ρ is independent on λ, the first term in Equation (2.148) can be reformulated in terms of the two integration limits

$$\begin{aligned} \int_0^1 \langle \Psi^\lambda | \frac{dV_{ext}^\lambda}{d\lambda} | \Psi^\lambda \rangle \, d\lambda &= \int \rho(\vec{r}) \left[\int_0^1 \frac{dV_{ext}^\lambda(\vec{r})}{d\lambda} d\lambda \right] d\vec{r} \\ &= \int \rho(\vec{r}) \left[V_{ext}^{\lambda=1}(\vec{r}) - V_{ext}^{\lambda=0}(\vec{r}) \right] d\vec{r} \\ &= \int \rho(\vec{r}) V_{ext}^{\lambda=1}(\vec{r}) d\vec{r} - \int \rho(\vec{r}) V_{ext}^{\lambda=0}(\vec{r}) d\vec{r} \end{aligned} \tag{2.149}$$

Considering that the energy of the non-interacting system is given by (Equation (2.121))

$$E^{\lambda=0} = \langle \Psi^{\lambda=0} | T | \Psi^{\lambda=0} \rangle + \int \rho(\vec{r}) V_{ext}^{\lambda=0}(\vec{r}) d\vec{r} \tag{2.150}$$

the expression for the energy of the fully interaction system reads

$$E^{\lambda=1} = \langle \Psi^{\lambda=0} | T | \Psi^{\lambda=0} \rangle + \int \rho(\vec{r}) V_{ext}^{\lambda=1}(\vec{r}) d\vec{r} + \int_0^1 \langle \Psi^\lambda | V_{ee} | \Psi^\lambda \rangle \, d\lambda \tag{2.151}$$

Since $V_{ext}^{\lambda=1}(\vec{r})$ represents the bare external potential, comparing Equation (2.151) with the same energy given in Equation (2.122) leads to

$$J[\rho] + E_{xc}[\rho] = \int_0^1 \langle \Psi^\lambda | V_{ee} | \Psi^\lambda \rangle \, d\lambda \tag{2.152}$$

This expression gives the relations between the exchange-correlation energy functional and the pair electron repulsion operator, that is $E_{xc}[\rho]$ can be obtained by integrating the electron-electron interaction over the coupling strength parameter λ and subtracting the classical Coulomb part $J[\rho]$.

By using the exchange-correlation hole as introduced in Equation (2.142), the argument of the integral on the right hand side of Equation (2.152) can be written as

$$\begin{aligned}\langle\Psi^\lambda|\, V_{ee}\,|\Psi^\lambda\rangle\, d\lambda &= \frac{1}{2}\int \frac{\rho_2^\lambda(\vec{r}_1,\vec{r}_2)}{|\vec{r}_1-\vec{r}_2|}d\vec{r}_1 d\vec{r}_2 \\ &= \frac{1}{2}\int \frac{\rho_1(\vec{r}_1)\rho_1(\vec{r}_2)}{|\vec{r}_1-\vec{r}_2|}d\vec{r}_1 d\vec{r}_2 + \frac{1}{2}\int \frac{\rho_1(\vec{r}_1)h_{xc}^\lambda(\vec{r}_1,\vec{r}_2)}{|\vec{r}_1-\vec{r}_2|}d\vec{r}_1 d\vec{r}_2 \\ &= J[\rho] + \frac{1}{2}\int \frac{\rho_1(\vec{r}_1)h_{xc}^\lambda(\vec{r}_1,\vec{r}_2)}{|\vec{r}_1-\vec{r}_2|}d\vec{r}_1 d\vec{r}_2 \end{aligned} \tag{2.153}$$

Comparing with Equation (2.152) allows to introduce the exchange-correlation hole potential $V_{xc}^\lambda(\vec{r})$ defined by the *adiabatic connection formula*

$$V_{xc}^\lambda(\vec{r}_1) = \frac{1}{2}\int \frac{h_{xc}^\lambda(\vec{r}_1,\vec{r}_2)}{|\vec{r}_1-\vec{r}_2|}d\vec{r}_2 \tag{2.154}$$

$$E_{xc} = \int_0^1 \langle\Psi^\lambda|\, V_{xc}^\lambda\,|\Psi^\lambda\rangle\, d\lambda = \int \rho(\vec{r})\left(\int_0^1 V_{xc}^\lambda(\vec{r})d\lambda\right)d\vec{r} \tag{2.155}$$

2.3.5 The RPA Correlation Energy in DFT Context

An approach for treating correlation effects in DFT is based on the idea to reformulate the potential energy of the electron-electron interaction in terms of density fluctuation around its average value. In this respect, the fluctuation-dissipation theorem (FDT) provides a link between density-density fluctuation and the response property (dissipation) of the system [12]. The FDT states that, at thermodynamic equilibrium, the response of a system to a small external perturbation is the same as its response to the spontaneous internal fluctuations without the perturbation. Specifically, the mean square fluctuation of a local one-particle observable in the ground state Ψ_0 can be related to the imaginary (dissipative) part of the density-density response function $\chi(\vec{r}_1,\vec{r}_2,\omega)$ [8]. This function describes the change of electronic density at $\vec{r}_1$ due to a change of the potential in $\vec{r}_2$ induced by an external field with frequency ω.

In the framework of the adiabatic connection approach, the exchange-correlation energy is obtained as a coupling-strength integrated exchange-correlation hole. The dissipation-fluctuation theorem allows to express the coupling strength integrand in terms of frequency-dependent density response function [4, 5]. Focusing on the correlation energy this leads to

$$E_c = -\frac{1}{2\pi}\int_0^1 d\lambda \int_0^\infty d\omega \int d\vec{r}_1 d\vec{r}_2 \frac{1}{|\vec{r}_1-\vec{r}_2|}\left[\chi_\lambda(\vec{r}_1,\vec{r}_2,i\omega) - \chi_0(\vec{r}_1,\vec{r}_2,i\omega)\right] \tag{2.156}$$

that represents the correlation energy within the adiabatic-connection fluctuation-dissipation (ACFD) theorem. In Equation (2.156) $\chi_\lambda(\vec{r}_1,\vec{r}_2,i\omega)$ and $\chi_0(\vec{r}_1,\vec{r}_2,i\omega)$ represent respectively the density-response functions for the λ-scaled and for the KS independent-particle systems. The latter is expressed in terms of KS orbitals ψ and orbitals energies ϵ

$$\chi_0(\vec{r}_1,\vec{r}_2,i\omega) = \sum_{ia}\left[\frac{\psi_i(\vec{r}_1)\psi_a(\vec{r}_1)\psi_i(\vec{r}_2)\psi_a(\vec{r}_2)}{\epsilon_i-\epsilon_a-i\omega} + \frac{\psi_i(\vec{r}_1)\psi_a(\vec{r}_1)\psi_i(\vec{r}_2)\psi_a(\vec{r}_2)}{\epsilon_i-\epsilon_a+i\omega}\right] \tag{2.157}$$

with the usual index convention for occupied and virtual orbitals. The ACDF formula in Equation (2.156) converts the problem of computing the correlation energy into one of computing the response function of a series of fictitious system along the adiabatic-connection path [12]. This is accomplished practically by introducing approximation for $\chi_\lambda(\vec{r}_1, \vec{r}_2, i\omega)$.

In this respect, the random-phase approximation is a possible approach for expressing the density-response function χ_λ. Within RPA, $\chi_\lambda^{\mathrm{RPA}}(\vec{r}_1, \vec{r}_2, i\omega)$ is obtained as the solution of the Dyson-type equation

$$\chi_\lambda^{\mathrm{RPA}}(\vec{r}_1, \vec{r}_2, i\omega) = \chi_0(\vec{r}_1, \vec{r}_2, i\omega) + \int d\vec{r} d\vec{r}' \chi_0(\vec{r}_1, \vec{r}, i\omega) \frac{\lambda}{|\vec{r} - \vec{r}'|} \chi_\lambda^{\mathrm{RPA}}(\vec{r}', \vec{r}_2, i\omega) \quad (2.158)$$

that eventually can be reduced to a finite dimensional linear algebra problem by introducing a basis [41].

More precisely, by representing the quantities in Equation (2.158) in the particle-hole basis $\{\psi_i(\vec{r}_1)\psi_a(\vec{r}_1)\}$, the RPA correlation energy can be obtained by the solution of the following eigenvalue problem [42]

$$\begin{pmatrix} \mathbf{A} & \mathbf{B} \\ -\mathbf{B} & -\mathbf{A} \end{pmatrix} \begin{pmatrix} X_i \\ Y_i \end{pmatrix} = \begin{pmatrix} X_i \\ Y_i \end{pmatrix} \omega_i \quad (2.159)$$

for which the $\mathbf{A}$ and $\mathbf{B}$ matrix elements are defined depending on the RPA approach considered. Within the direct-RPA (dRPA) approach, that is RPA without including exchange contributions [36, 43], the orbital rotation Hessian matrices $\mathbf{A}$ and $\mathbf{B}$ are defined as:

$$(A - B)_{ia,jb} = (\epsilon_a - \epsilon_i)\delta_{ij}\delta_{ab} \quad (2.160)$$

$$(A + B)_{ia,jb} = (\epsilon_a - \epsilon_i)\delta_{ij}\delta_{ab} + 2(ia|jb) \quad (2.161)$$

where all matrices have dimension $ov \times ov$ being o and v the number of occupied and virtual orbital respectively. This allows to reformulate Equation (2.159) into a Hermitian problem

$$\mathbf{MZ} = \mathbf{Z}\boldsymbol{\omega}^{\mathbf{2}}, \quad \mathbf{Z}^{\mathbf{T}}\mathbf{Z} = \mathbf{1} \quad (2.162)$$

with

$$\mathbf{M} = (\mathbf{A} - \mathbf{B})^{1/2}(\mathbf{A} + \mathbf{B})(\mathbf{A} - \mathbf{B})^{1/2}. \quad (2.163)$$

After solving Equation (2.162), the dRPA correlation energy is given by [35]

$$E_c^{\mathrm{dRPA}} = \frac{1}{2}\mathrm{Tr}(\omega - \mathbf{A}) = \frac{1}{2}\mathrm{Tr}(\mathbf{M}^{1/2} - \mathbf{A}) \quad (2.164)$$

Finally, in the RPA correlation energy framework, the exchange energy is accounted in terms of exact exchange (EXX) evaluated using the ground state wavefunction of the non-interacting system ($\lambda = 0$)

$$E_x = \langle \Psi^{\lambda=0} | \, V_{ee} \, | \Psi^{\lambda=0} \rangle - J[\rho] = -\frac{1}{2}\sum_{ij}(ij|ij) \quad (2.165)$$

Since for the non-interacting system Ψ is a single Slater determinant, the exchange energy takes the same form as that obtained in the Hartree-Fock case. However the value differs in practice due to the deviations between the KS and HF orbitals that enter Equation (2.165).

The exchange and correlation energy given above define the (EXX+RPA) method, that can be ultimately considered as a post Kohn-Sham approach. In fact, the ground state single-particle wave functions and orbital energies obtained from a given KS-DFT procedure, are subsequently used as input to compute the EXX energy and the RPA correlation energy. For this reason the energy obtained in this way is often referred as (EXX+RPA)@DFT, where the DFT label denotes the density functional approximation employed for solving the KS equations.

Reported here are only the main features of the RPA correlation energy method, for more details see the references [9, 10, 12, 42].

Chapter 3

MP2 Energy in the Condensed Phase: An Efficient and Massively Parallel Gaussian and Plane Waves Approach [44]

A novel algorithm, based on a hybrid Gaussian and Plane Waves (GPW) approach, is developed for the canonical second-order Møller-Plesset perturbation energy (MP2) of finite and extended system. The key aspect of the method is that the electron repulsion integrals $(ia|\lambda\sigma)$ are computed by direct integration between the products of Gaussian basis functions $\lambda\sigma$ and the electrostatic potential arising from a given occupied-virtual pair density ia. The electrostatic potential is obtained in a plane waves basis set after solving the Poisson equation in Fourier space. In particular for condensed phase systems, this scheme is highly efficient. Furthermore, our implementation has low memory requirements and displays excellent parallel scalability up to 100 000 processes. In this way, canonical MP2 calculations for condensed phase systems containing hundreds of atoms or more than 5000 basis functions can be performed within minutes. Lattice constants and cohesive energies of various molecular crystals have been studied with MP2 and double hybrid functionals.

3.1 Introduction

The second-order Møller-Plesset (MP2) energy is an effective correction to the Hartree-Fock ground state energy that accounts for electron correlation effects. It is obtained from Rayleigh-Schrödinger perturbation theory, for which the zeroth order Hamiltonian is chosen as the sum of the one electron Fock operators [13, 14]. MP2 is appealing because it recovers a relatively large part of the dynamic correlation, and maintains an easy and compact formulation. Most notably, MP2 introduces dispersion, which is an essential non-covalent interaction. Furthermore, MP2 is also available for condensed phase systems, i.e. including periodic boundary conditions [45–51]. With the introduction of Double-Hybrid Density Functionals [17, 19], MP2-like correlation has also established itself in Density Functional Theory (DFT). In Double-Hybrids, an MP2-like term obtained from

the Kohn-Sham orbitals and eigenvalues is mixed into the correlation energy. However, the advantages of MP2 come at a computational cost that is high compared to that of Hartree-Fock or traditional DFT. In its canonical formulation, MP2 scales a $O(n^5)$, where n represents the number of basis functions, and a large amount of memory is needed to store the intermediates of the calculation. Furthermore, MP2 calculations need larger basis sets than DFT to reach a similar convergence. In order to extend the applicability of MP2 to large system these limitations have to be overcome [52].

Various reformulations of the MP2 energy expression, and new algorithms, have been proposed to address these limitations. Reducing the formal $O(n^5)$ scaling is achieved with methods such as local MP2 [49, 53–61] (LMP2) and Laplace-Transformed MP2 [48, 62–67]. The prefactor of the various terms that dominate for smaller systems can be reduced with the resolution of identity approximation [46, 47, 68–73] (RI-MP2), while explicitly correlated methods speedup the convergence of the MP2 energy with respect to basis set size[74] (F12-MP2). Despite this progress, calculations with good basis sets on systems containing fifty or more heavy atoms remain computationally demanding with MP2 or double hybrid DFT. In order to perform such calculations with acceptable time to solution, massively parallel computing is an indispensable tool. A variety of MP2 algorithms suitable for parallel architectures has been proposed[75–85] and these algorithms have demonstrated good scalability up to a few hundred cores.

Here, we present a novel MP2 algorithm that is particularly suitable for the condensed phase and has been designed to achieve excellent scalability on modern massively parallel architectures having 1000s-100000s of cores. The prefactor of the $O(n^5)$ term is minimal and the memory usage per core is small. It is based on the Gaussian and Plane Wave (GPW) approach[86], which allows for avoiding the computation of four center electron repulsion integrals (ERI) over Gaussian basis functions $(\mu\nu|\lambda\sigma)$. In conventional canonical MP2 algorithms, the computation of these integrals and their transformation into the molecular orbital basis is usually the most time-consuming step. This step is furthermore difficult to parallelize efficiently, involving significant communication and difficult load balancing issues. In the Gaussian and Plane Wave MP2 (GPW-MP2) method, half transformed ERIs of the type $(ia|\lambda\sigma)$ are directly computed in a communication free way. This is achieved by the direct computation of the electrostatic potential of the pair density $\rho^{ia} = \psi_i \cdot \psi_a$ in an auxiliary plane waves basis by means of Fast Fourier Transforms (FFTs) and the numerical integration of this potential in real space over products of pairs of Gaussian basis functions $\lambda\sigma$. With this strategy, only fully transformed ERIs $(ia|jb)$ are communicated for the calculation of the exchange like part of the MP2 energy. The efficiency derives from the use of regular auxiliary grids and FFTs, which distinguishes the method from other approaches employing numerical integration [87, 88]. We report parallel scalability up to 100 000 processes with 80% efficiency, allowing calculations on molecular crystals containing more than 5000 basis functions within minutes. We further validate the GPW-MP2 method by performing calculations on molecular crystals with extended basis sets.

3.2 The Gaussian and Plane Wave MP2 method

In the canonical orbital formalism, the closed shell MP2 correlation energy $E^{(2)}$ is obtained as

$$E^{(2)} = -\sum_{ij,ab}^{occ,vir} \frac{(ia|jb)[2(ia|jb) - (ib|ja)]}{\epsilon_a + \epsilon_b - \epsilon_i - \epsilon_j}. \tag{3.1}$$

Indices i, j refer to occupied and a, b to virtual canonical orbitals, and ϵ_p to the corresponding orbital energy. The ERIs over molecular orbitals (MO ERI) are given by

$$(ia|jb) = \int\int \psi_i(\vec{r}_1)\psi_a(\vec{r}_1)\frac{1}{\vec{r}_{12}}\psi_j(\vec{r}_2)\psi_b(\vec{r}_2)d\vec{r}_1 d\vec{r}_2 \tag{3.2}$$

and conventionally computed by a four index transformation from ERIs over atomic orbitals (AO ERI) $(\mu\nu|\lambda\sigma)$ as

$$(ia|jb) = \sum_{\mu\nu\lambda\sigma} (\mu\nu|\lambda\sigma)C_{\mu i}C_{\nu a}C_{\lambda j}C_{\sigma b} \tag{3.3}$$

where $C_{\mu i}$ represent elements of the MOs coefficient matrix and Greek indices refer to AOs, a linear combination of Gaussian basis functions in our approach. For systems described by periodic boundary conditions (PBC), Brillouin sampling is implicitly implied for Equation (3.1), but here we will assume that Γ-point sampling is sufficient for systems with a sufficiently large unit cell and band gap [45, 50, 89]. In the periodic case, AOs and the integrals in Equation (3.2) must take the PBC into account [86].

The basis of the Gaussian and Plane Wave MP2 method (GPW-MP2) is the direct formulation of the half transformed ERIs based on the electrostatic potential v^{ia} of the pair density ρ^{ia}

$$\begin{aligned}
(ia|\lambda\sigma) &= \int\int \psi_i(\vec{r}_1)\psi_a(\vec{r}_1)\frac{1}{\vec{r}_{12}}\phi_\lambda(\vec{r}_2)\phi_\sigma(\vec{r}_2)d\vec{r}_1 d\vec{r}_2 \\
&= \int \left[\int \frac{\rho^{ia}(\vec{r}_1)}{\vec{r}_{12}} d\vec{r}_1\right] \phi_\lambda(\vec{r}_2)\phi_\sigma(\vec{r}_2)d\vec{r}_2 \\
&= \int v^{ia}(\vec{r}_2)\phi_\lambda(\vec{r}_2)\phi_\sigma(\vec{r}_2)d\vec{r}_2
\end{aligned} \tag{3.4}$$

The form of the last equation is essentially identical to the one used in the GPW method[86] to compute matrix elements of the Hartree potential. Thus, the highly efficient implementation of that operation in CP2K[90] can be directly used and we refer to Ref. [91] for a detailed discussion.

Central in the GPW method is the representation of the density $\rho^{ia}(\vec{r})$ on a regular grid, which can be considered equivalent to an expansion of the density in an auxiliary basis of plane waves (PW). The expansion is given by

$$\rho^{ia}(\vec{r}) \approx \frac{1}{\Omega}\sum_{|\vec{G}|\leq G_c} \rho^{ia}(\vec{G})e^{i\vec{G}\cdot\vec{r}} \tag{3.5}$$

where the sum over the reciprocal lattice vectors $\vec{G}$ is determined by the resolution of the grid. $\rho^{ia}(\vec{G})$ are the Fourier coefficients of the density, and Ω is the volume of the

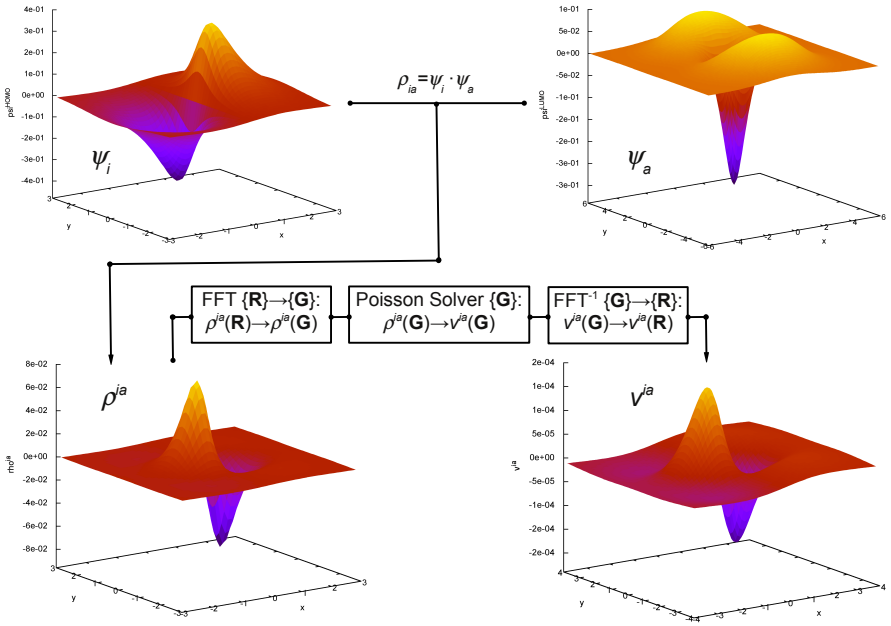


Figure 3.1: Schematic representation of the operations involved in the calculation of the potential $v^{ia}(\vec{R})$ starting from the occupied $\psi_i(\vec{R})$ and virtual $\psi_a(\vec{R})$ orbitals. Each of the four plots reports the value of $\psi_i(\vec{R})$, $\psi_a(\vec{R})$, $\rho^{ia}(\vec{R})$ and $v^{ia}(\vec{R})$ in the xy plane in the case of the ψ_i = HOMO and ψ_a = LUMO of H_2O (oriented with molecular plane orthogonal to the xy plane and parallel to the yz plane).

simulation cell. Conventionally, the resolution of the grid is specified as the energy cutoff $\frac{1}{2}G_c^2$ that limits the kinetic energy of the PWs. Fast Fourier transforms (FFTs) efficiently change representation between real space ($\rho^{ia}(\vec{R})$) and reciprocal space ($\rho^{ia}(\vec{G})$). In particular, for a grid with S grid points, the transformation can be performed in linear scaling time ($O(S \log S)$). In reciprocal space, it becomes straightforward to solve the Poisson equation for the potential v^{ia}

$$v^{ia}(\vec{G}) = \frac{4\pi}{G^2}\rho^{ia}(\vec{G}) \tag{3.6}$$

and an additional back FFT (FFT^{-1}) will yield the potential in real space. The orthonormality of the orbitals implies that $\rho^{ia}(\vec{G}=0)=0$ and divergence at $\vec{G}=0$ is thus avoided [92]. Note that the PW auxiliary basis is a natural choice for periodic systems, but it can equally be used for gas phase or surface calculations. Indeed, once the density is specified on a regular grid efficient methods are available for solving the Poisson equation with free (for example cluster or slab) boundary conditions[93–95].

All the operations previously described, that allows for the computation of the electrostatic potential $v^{ia}(\vec{R})$ starting from the wave function's $\psi_i(\vec{R})$ and $\psi_a(\vec{R})$, are schematically depicted in Figure 3.1 in the case of the frontier orbitals of the water molecule.

The simplicity of the GPW method has as a drawback that all-electron calculations are not possible, and that pseudopotentials have to be employed in order to have densities that are smooth. The Gaussian and Augmented Plane Wave (GAPW) scheme [96, 97] overcomes this limitation and is suitable for all-electron calculations. However, whereas this method is available in CP2K for all-electron DFT calculations, our MP2 implementation

is currently limited to the GPW method only.

Once the potential v^{ia} is available on a regular real space grid, the numerical integration over the basis functions is performed by summing the product of the value of the potential and the primitive Gaussian functions (PGFs) over the grid points. Within a given threshold (ϵ_{grid}), all non-zero matrix elements for a given pair ia can be obtained in linear scaling time. This is possible since only pairs of overlapping Gaussians need to be considered, and only a finite number of grid points within a spherical region around the center of the PGF is required. A further gain in efficiency is obtained by employing a multi-grid technique that represents the potential v^{ia} on grids with increasingly coarser grid spacing. Depending on the smoothness or width of the PGF, the appropriate grid is selected so that the number of points employed for the integration is essentially independent of the exponent of the PGF. The accuracy of the multi-grid scheme is fixed by specifying a relative cutoff (E_{cut}^{rel}) that specifies the E_{cut} of the grid that will be employed for a PGF with exponent 1.0.

Finally, $(ia|\lambda\sigma)$ integrals are transformed into MO ERIs using sparse matrix multiplication. Introducing for a given pair ia the matrix of half transformed ERIs $\mathbf{B}^{ia}$ ($(ia|\lambda\sigma) = B^{ia}_{\lambda\sigma}$), the matrix of MO ERIs $\mathbf{V}^{ia}$ is obtained by two index transformations as $\mathbf{V}^{ia} = \mathbf{C}_o^\dagger \mathbf{B}^{ia} \mathbf{C}_v$, where $\mathbf{C}_o$ and $\mathbf{C}_v$ represent the coefficient matrices of the occupied and virtual orbitals. The multiplication by $\mathbf{C}_v$ can exploit the sparsity of $\mathbf{B}^{ia}$, implying an $O(nv)$ scaling per ia pair, while the final multiplication cannot exploit sparsity and is asymptotically dominant, scaling as $O(onv)$. o, v, and n refer to the number of occupied, virtual, and total orbitals respectively. The thresholding in the sparse matrix multiplication is enforced using a threshold $\epsilon_{filter} \approx \epsilon_{grid}$. As we will show below, the overall accuracy of the MP2 energy can be well controlled, and for the systems tested here, is on the order of $10^{-7} - 10^{-8}$ a.u. per heavy atom for $E_{cut} = 300$ Ry, $E_{cut}^{rel} = 50$ Ry, $\epsilon_{filter} = \epsilon_{grid} = 10^{-8}$. E_{cut} depends on the largest exponent of the basis used, while the other parameters are system independent.

3.3 Implementation of the Gaussian and Plane Wave MP2 method

3.3.1 Serial algorithm

The pseudocode for the serial algorithm is presented in Figure 3.2. In a first step, the wavefunctions of all occupied orbitals ($\psi_i(\vec{R})$) are precalculated on real space grids, which speeds up the calculation of the pair density ρ^{ia} in the main loop but is not strictly necessary. In the next step, the outer loop iterates over all virtual orbitals a, while the inner one loops over occupied orbitals i. For each value of a, the MO ERIs $(ia|jb)$ for all other indices are available after the inner loop, making it possible to compute both Coulomb and exchange contributions to $E^{(2)}$ with $O(o^2v)$ memory usage, which compares favorably with the $O(on^2)$ memory required in the standard direct canonical algorithm. Other objects, such as the stored grids and molecular orbital coefficients, require memory scaling no worse than quadratic with system size. In the inner loop, all the grid operations and integral transformations for a given pair ia are performed. The density $\rho^{ia}(\vec{R})$ on the real space grid is simply obtained by multiplying the values for $\psi_i(\vec{R})$ and $\psi_a(\vec{R})$, and

Loop over all i occupied orbitals

Calculate wavefunction $\psi_i(\vec{R})$ on the real space grid

Store $\psi_i(\vec{R})$

End i Loop

Loop over all a virtual orbitals

Calculate wavefunction $\psi_a(\vec{R})$ on the real space grid

Loop over all i occupied orbitals

Compute $\rho^{ia}(\vec{R}) = \psi_i(\vec{R}) * \psi_a(\vec{R})$ on the real space grid

Transfer $\rho^{ia}(\vec{R}) \rightarrow \rho^{ia}(\vec{G})$: $\rho^{ia}(\vec{G}) = \mathbf{FFT}[\rho^{ia}(\vec{R})]$

Solve the Poisson equation: $\rho^{ia}(\vec{G}) \rightarrow v^{ia}(\vec{G})$

Transfer $v^{ia}(\vec{G}) \rightarrow v^{ia}(\vec{R})$: $v^{ia}(\vec{R}) = \mathbf{FFT^{-1}}[v^{ia}(\vec{G})]$

Integrate potential in real space: $(ia|\lambda\sigma) = \mathrm{B}^{ia}_{\lambda\sigma} = \int v^{ia}(\vec{R})\phi_\lambda(\vec{R})\phi_\sigma(\vec{R})d\vec{R}$

Index transformations $\mathbf{V}^{ia} = \mathbf{C}_o^\dagger(\mathbf{B}^{ia}\mathbf{C}_v)$

Store $\mathbf{V}^{ia}$

End i Loop

$E^{(2)} = E^{(2)} + \sum_{ij,b} \frac{(ia|jb)[2(ia|jb)-(ja|ib)]}{\epsilon_i+\epsilon_j-\epsilon_a-\epsilon_b}$

End a loop

Figure 3.2: Pseudocode for the serial implementation of the GPW-MP2 energy.

is transformed by **FFT** to reciprocal space, where the Poisson equation is solved. An inverse FFT yields the potential in real space, which is used in the numerical integration procedure to yield the $(ia|\lambda\sigma)$ integrals. The latter are stored as a sparse matrix, which can directly be employed in the following index transformations that yield the MO ERIs $(ia|jb)$. Note again that the first transformation can exploit the sparsity of the $\mathbf{B}^{ia}$ matrix, while the second transformation cannot. The main features of the serial GPW-MP2 algorithm are summarized in Table 3.1.

3.3.2 Parallel algorithm

The parallel algorithm for the GPW-MP2 energy calculation has been designed to enable calculations on large systems and to display excellent scalability. This implies that the computational load and the amount of data communicated per process decreases linearly as the number of processes (N_p) increases. Furthermore, the computational load is easily balanced and the communication pattern involves relatively large messages between a small subset of all possible pairs of processes. Additionally, no significant data is fully replicated and memory usage can be decreased as the number of processes is increased. This is achieved by a multi-level parallelization scheme, and a careful process layout. The first level of parallelization corresponds to the work performed for a given ia pair. The second level of parallelization corresponds to a distribution of the nearly independent calculations for each of the ia pairs. The N_p processes are therefor split in N_G groups, each group working on a given ia pair and consisting of N_w processes ($N_p = N_G N_w$). The first level of parallelization is complicated, involving parallel FFTs, halo-exchanges, and

Assign each process its coordinate-triplet (n_i, n_a, n_w)

Create ranges $[i^{n_i}_{start}, i^{n_i}_{end}]$, $[a^{n_a}_{start}, a^{n_a}_{end}]$, $[b^{n_w}_{start}, b^{n_w}_{end}]$

Loop over i occupied orbitals $(i^{n_i}_{start} \leq i \leq i^{n_i}_{end})$

Calculate wavefunction $\psi_i(\vec{R})$ on the real space grid

Store $\psi_i(\vec{R})$

End i Loop

Loop over a virtual orbitals $(a^{n_a}_{start} \leq a \leq a^{n_a}_{end})$

Calculate wavefunction $\psi_a(\vec{R})$ on the real space grid

Loop over i occupied orbitals $(i^{n_i}_{start} \leq i \leq i^{n_i}_{end})$

Compute $\rho^{ia}(\vec{R}) = \psi_i(\vec{R}) * \psi_a(\vec{R})$ on the real space grid

Transfer $\rho^{ia}(\vec{R}) \rightarrow \rho^{ia}(\vec{G})$: $\rho^{ia}(\vec{G}) = \mathbf{FFT}[\rho^{ia}(\vec{R})]$

Solve Poisson's Equation: $\rho^{ia}(\vec{G}) \rightarrow v^{ia}(\vec{G})$

Transfer $v^{ia}(\vec{G}) \rightarrow v^{ia}(\vec{R})$: $v^{ia}(\vec{R}) = \mathbf{FFT^{-1}}[v^{ia}(\vec{G})]$

Integrate Potential in real space: $(ia|\lambda\sigma) = \mathrm{B}^{ia}_{\lambda\sigma} = \int v^{ia}(\vec{R})\phi_\lambda(\vec{R})\phi_\sigma(\vec{R})d\vec{R}$

Index transformation $\mathbf{V}^{ia} = \mathbf{C}_o^\dagger(\mathbf{B}^{ia}\mathbf{C}_v)$

Redistribute and store $\mathbf{V}^{ia}$ (all j, $b^{n_w}_{start} \leq b \leq b^{n_w}_{end}$)

End i Loop

Loop over N_i processes with same n_a and n_w but different n_i

determine the coordinates n_i^S and n_i^R of sending and receiving process.

Receive $(ja|ib)$ from (n_i^R, n_a, n_w)

$(i^{n_i^R}_{start} \leq j \leq i^{n_i^R}_{end}, i^{n_i}_{start} \leq i \leq i^{n_i}_{end}, b^{n_w}_{start} \leq b \leq b^{n_w}_{end})$

Send $(ia|jb)$ to (n_i^S, n_a, n_w)

$(i^{n_i^S}_{start} \leq j \leq i^{n_i^S}_{end}, i^{n_i}_{start} \leq i \leq i^{n_i}_{end}, b^{n_w}_{start} \leq b \leq b^{n_w}_{end})$

$E^{(2)} = E^{(2)} + \sum_{ij,b} \frac{(ia|jb)[2(ia|jb)-(ja|ib)]}{\epsilon_i+\epsilon_j-\epsilon_a-\epsilon_b}$

$(i^{n_i^R}_{start} \leq j \leq i^{n_i^R}_{end}, i^{n_i}_{start} \leq i \leq i^{n_i}_{end}, b^{n_w}_{start} \leq b \leq b^{n_w}_{end})$

End loop over processes

End a loop

Global summation of $E^{(2)}$

Figure 3.3: Pseudocode of the parallel GPW-MP2 energy.

Table 3.1: Analysis of the computational complexity of the serial and parallel GPW-MP2 energy algorithms as a function of the calculation parameters: n number of basis functions, o and v number of occupied and virtual orbitals, S grid size, N_i and N_a number of occupied and virtual divisions, N_G and N_w number of groups and group size, N_p number of processes. N_i, N_a, N_G, N_w and N_p are related by $N_G = N_i N_a$ and $N_p = N_G N_w$.

	Serial		Parallel	
	Memory	Execution Time	Memory	Execution Time
Wave functions calculation	oS	nS	$\frac{oS}{N_i N_w}$	$\left(\frac{o}{N_i} + \frac{v}{N_a}\right)\frac{S}{N_w}$
FFT and Poisson solver	S	$ovS \log S$	$\frac{S}{N_w}$	$\frac{ovS \log S}{N_p}$
v^{ia} integration	S	ovn	$\frac{S}{N_w}$	$\frac{ovn}{N_p}$
1^{st} quarter transformation	on	o^2vn	$\frac{on}{N_w}$	$\frac{o^2vn}{N_p}$
2^{nd} quarter transformation	ov	o^2v^2n	$\frac{ov}{N_w}$	$\frac{o^2v^2n}{N_p}$
MP2 energy contraction	o^2v	o^2v^2	$\frac{o^2v}{N_i N_w}$	$\frac{o^2v^2}{N_p}$
Communication	-	-	$\frac{o^2v}{N_i^2 N_w}$	$\frac{N_i - 1}{N_i}\frac{o^2v^2}{N_p}$

sparse matrix multiplications over N_w processes. However, this level is readily available, as it corresponds to the standard parallelization scheme for DFT calculations in CP2K [91]. For example, for the 3rd and 4th quarter transformation, both molecular orbitals and half transformed integrals are represented as sparse distributed matrices, which can be multiplied using a generally purpose parallel sparse matrix matrix multiplication library we have recently developed and use extensively in linear scaling calculations [98, 99]. As a rule of thumb, reasonable speedups are observed as long as $N_w \leq o$, while memory usage benefits from the nearly perfect distribution of the grids and sparse matrices. Nevertheless, the best performance is obtained if groups do not communicate across nodes, ideally, if memory permits, $N_w = 1$. The second level of parallelization is more straightforward, as it only requires inter-process communications of fully transformed ERIs. In order to distribute the ia pairs efficiently, the occupied orbitals i and the virtual orbitals a are split into N_i and N_a disjoint ranges respectively. A 2D Cartesian layout with dimensions $N_i \times N_a$ is considered for the N_G groups ($N_G = N_i N_a$) giving each group coordinates (n_i, n_a) and the corresponding index ranges $[i_{start}^{n_i}, i_{end}^{n_i}]$ and $[a_{start}^{n_a}, a_{end}^{n_a}]$. Additionally, each of the N_w processes within a group is given an index n_w, so that a processes is uniquely identified by its coordinate triplet (n_i, n_a, n_w). Finally, the b index is split in N_w ranges $[b_{start}^{n_w}, b_{end}^{n_w}]$, while a splitting of j is not necessary.

The pseudo-code of the parallel algorithm is shown in Figure 3.3 and it follows the serial algorithm closely (but with restricted index ranges for i and a), until the end of the inner loop over i. Here, as a last step of the inner loop, the matrix of fully transformed integrals $\mathbf{V}^{ia}$ is redistributed within the group, so that the full range of j and the restricted

range of b, corresponding the process' index n_w, is stored locally. After completion of the loop over i the inter-group communication takes place, and here the benefit of the process layout becomes apparent. Indeed, only processes with identical coordinates n_a and n_w need to exchange data, i.e. only within small subgroups of size N_i communication takes place. This is due to the fact that for a given pair ab, only $(ia|jb)$ and $(ja|ib)$ need to be simultaneously available to a process. Since each process stores only ERIs $(ia|jb)$ for i in $[i_{start}^{n_i}, i_{end}^{n_i}]$ and all j, the locally held integrals can be contracted if integrals are received from all the other processes that store the ERIs for which $i \notin [i_{start}^{n_i}, i_{end}^{n_i}]$ for the current ab pair. This communication step is easily accomplished by employing the standard message passing interface (MPI) point-to-point communication protocol [100]. The fact that the computational effort for every ia pair is essentially the same, implies that processes arrive well synchronized at the communication step, contributing to the scalability of the algorithm. The size of each message send is $O\left(\frac{vo^2}{wN_i^2}\right)$, while the number of message exchanges, including the loop over a, is $O\left((N_i - 1)\frac{v}{N_a}\right)$, yielding an expected communication time $O\left(\frac{N_i - 1}{N_i}\frac{o^2v^2}{p}\right)$ in the bandwidth-limited regime. Note that if only the spin opposite (SO) component of the MP2 energy is required, the MO ERIs do not need to be communicated among processes, yielding an essentially communication free algorithm with reduced $O(on/N_w)$ memory usage. The main features of the parallel GPW-MP2 algorithm are summarized in Table 3.1.

3.4 Benchmark Calculations

3.4.1 Computational Details

Basis Sets, thresholds and pseudopotentials

The GPW-MP2 method as implemented in CP2K[90] has been employed for all calculations in this manuscript. The MP2 calculations are based on pseudopotentials of the form suggested by Goedecker, Teter and Hutter (GTH) in Ref. [101] but specifically parameterized for Hartree-Fock (HF) calculations. In this way, core states do not need to be represented and valence orbitals are smooth, as required by the GPW method. Valence-only basis sets have been generated for use with these pseudopotentials, and are suitable for MP2, i.e. of the correlation consistent type. The basis sets have been labeled as cc-DZVP, cc-TZVP and cc-QZVP, denoting increasing quality. The number of primitive Gaussian functions (for the valence only) has been chosen to be 4, 5 and 6 for cc-DZVP, cc-TZVP and cc-QZVP respectively. The three primitive Gaussian functions with higher exponent have been used to generate a contracted Gaussian, while the others have been kept uncontracted as in the split valence scheme. For the cc-QZVP basis the most diffuse functions have exponents H=0.13906, Li=0.16636, C=0.0597, O=0.10700. The basis sets have been augmented by the polarization functions taken from the all-electron basis set cc-pVXZ (X=D,T,Q) of Dunning [102, 103], up to g-functions. For further details about the basis and pseudopotential parameters we refer to the supporting information of Ref [44]. Despite the relatively diffuse nature of the Gaussian primitives, our robust

implementation of Hartee-Fock exchange[104, 105] allows for stable calculations in the condensed phase [105, 106]. The Schwarz screening threshold for the HF calculations is in the range $10^{-8} - 10^{-10}$ for the molecular crystals, going down to 10^{-14} for bulk LiH with the most extended basis set. Periodic calculations are based on a truncated Coulomb operator for Hartree-Fock exchange[105], using approximately half the length of the smallest edge of the simulation cell as truncation radius. The threshold for the SCF convergence was 10^{-7} or tighter. The PW cutoff for the HF or DFT part of the calculations was $E_{cut} = 1200Ry$ to guarantee convergence of the exchange and correlation functional, at small cost compared to the MP2 calculation. The MP2 calculation employed high quality PW cutoffs of $E_{cut} = 300$ Ry, $E_{cut}^{rel} = 50$ Ry, $\epsilon_{filter} = 10^{-7}$, and $\epsilon_{grid} = 10^{-6}$, unless mentioned otherwise. Gas phase systems have been computed using cluster boundary conditions for solving the Poisson equation [95].

Geometries

For all molecular crystals, supercells have been generated by replicating the unit cell, so that the smallest edge was larger than 9Å, in order for the Γ-point approximation to be reasonable. The geometries of the NH_3 and CO_2 crystals have been built from the experimental lattice parameter and the space group ($a = 5.048$ Å , $P2_13$, for NH_3; $a = 5.55$ Å , Pa-3, for CO_2) as detailed in Ref. [107]. The experimental geometries of the other molecular crystals have been retrieved from the Cambridge Structural Database (CSD)[108]. These positions of the hydrogen atoms of these geometries have been further relaxed at the DFT/BLYP [109, 110] level employing the cc-TZVP basis set. The main features of the structure of each crystal, together with the supercell used in the calculation and the CSD refcode, are reported in the caption of Figure 3.4. The LiH crystal geometry is based on the experimental value of the lattice parameter ($a = 4.084$ Å).

Cohesive Energies and Lattice Parameter Optimization

The counterpoise (CP) corrected cohesive energy per molecule at a given volume V has been computed as

$$E_{coh}^{CP}(V) = \frac{E_{supercell}(V)}{N_{mol}} - E_{mol}^{gas} - E_{mol+ghost}^{crystal}(V) + E_{mol}^{crystal}(V). \tag{3.7}$$

Here, N_{mol} is the number of molecules per supercell, $E_{supercell}(V)$ the total energy of the supercell, and $E_{mol+ghost}^{crystal}(V)$, $E_{mol}^{crystal}(V)$, and E_{mol}^{gas} the total energy of an isolated molecule in either the crystal geometry ($E_{mol+ghost}^{crystal}(V)$ and $E_{mol}^{crystal}(V)$) or a gas phase geometry (E_{mol}^{gas}). $E_{mol+ghost}^{crystal}(V)$, includes ghost atoms from the 12 nearest neighbor molecules in the case of NH_3 and CO_2, while only the first coordination shell has been retained for the other molecular crystals. This procedure has also been followed in Ref. [107] and Ref. [111]. In the case of LiH, this term has been computed using the fully periodic crystal geometry of the employed supercell. The gas phase geometry employed for NH_3 and CO_2 corresponds to the one specified in Ref. [107]. The other molecular crystals have been computed twice, once with gas phase geometries relaxed at the B3LYP/cc-TZVP level[110, 112, 113], and once, for direct comparison with Ref. [111], using the crystal geometry also for the gas phase geometry.

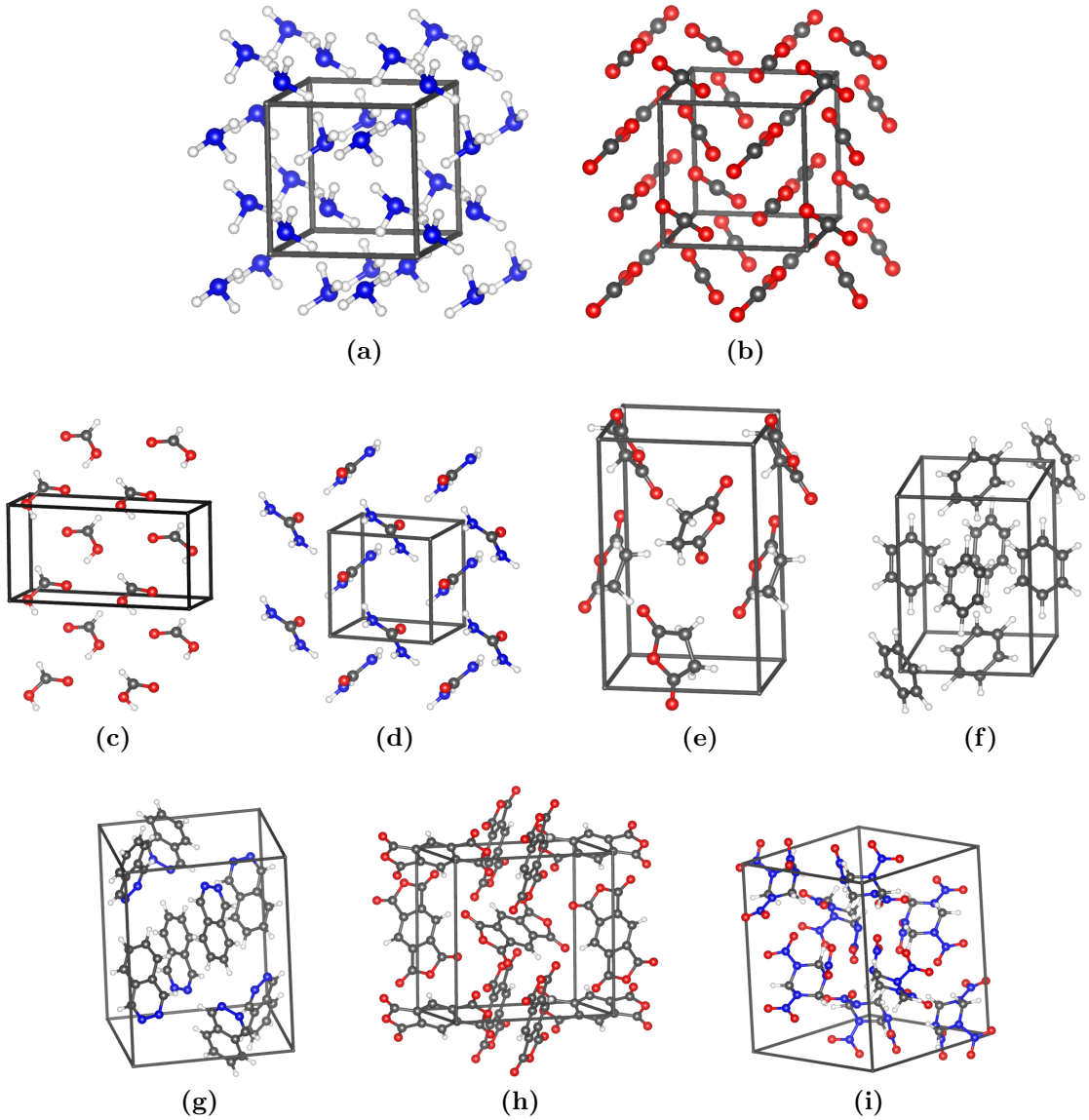


Figure 3.4: Unit cell and molecules of the supercell for the molecular crystals studied. For all crystals $\alpha = \beta = \gamma = 90°$, all lattice parameters expressed in Å. (a) NH_3, $a = b = c = 5.048$, supercell $2 \times 2 \times 2$, $N_{mol} = 32$ (b) CO_2, $a = b = c = 5.55$, supercell $2 \times 2 \times 2$, $N_{mol} = 32$ (c) Formic Acid, $a = 10.241$ $b = 3.544$ $c = 5.356$, supercell $1 \times 3 \times 2$ (refcode FORMAC01), $N_{mol} = 24$ (d) Urea, $a = b = 5.645$ $c = 4.704$, supercell $2 \times 2 \times 2$ (refcode UREAXX09), $N_{mol} = 16$ (e) Succinic Anhydride, $a = 5.4257$ $b = 6.9746$ $c = 11.7167$, supercell $2 \times 2 \times 1$ (refcode SUCANH15), $N_{mol} = 16$ (f) Benzene, $a = 7.398$ $b = 9.435$ $c = 6.778$, supercell $2 \times 1 \times 2$ (refcode BENZEN07), $N_{mol} = 16$ (g) 2,3-Diazanaphthalene, $a = 13.695$ $b = 10.557$ $c = 9.285$, supercell $1 \times 1 \times 1$ (refcode DAZNAP), $N_{mol} = 8$ (h) Pyromellitic Dianhydride, $a = b = 10.792$ $c = 7.4128$, supercell $1 \times 1 \times 2$ (refcode PYMDAN), $N_{mol} = 8$ (i) Trinitro-Triazacyclohexane, $a = 13.182$ $b = 11.574$ $c = 10.709$, supercell $1 \times 1 \times 1$ (refcode CTMTNA), $N_{mol} = 8$

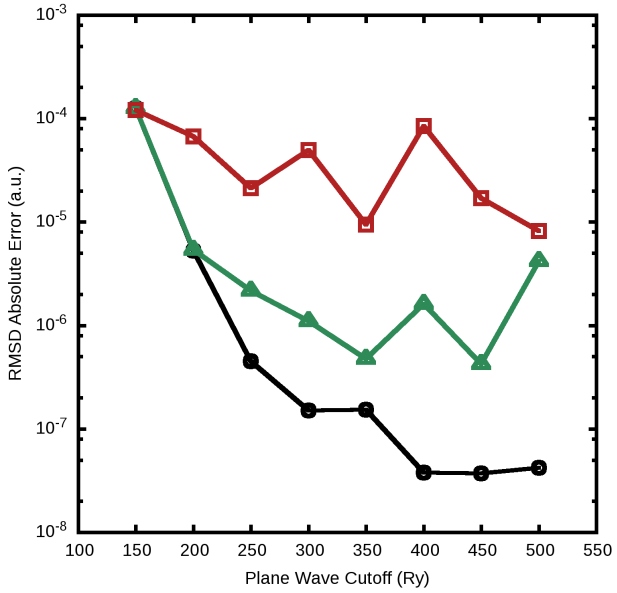
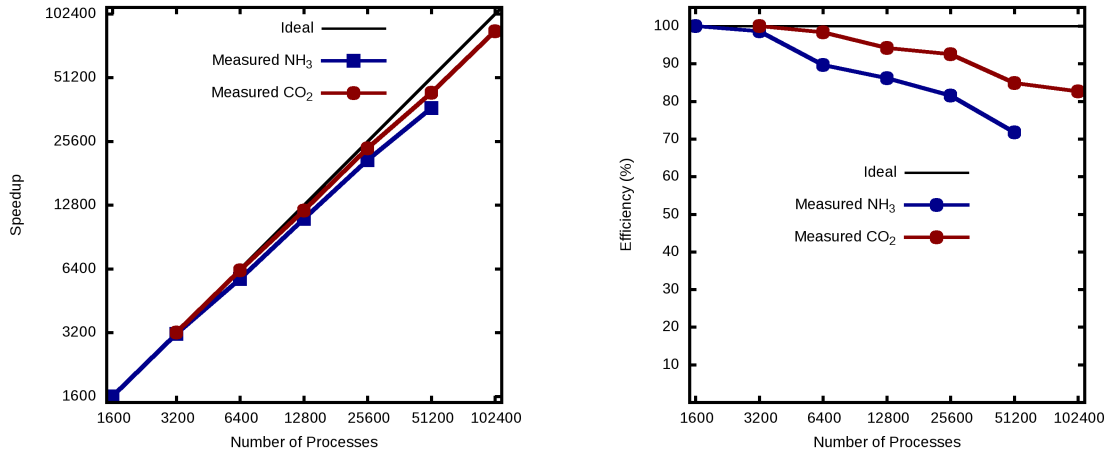


Figure 3.5: Root-mean-square deviation (RMSD) of the absolute error for the GPW-MP2 energy as a function of the plane wave cutoff (E_{cut}) for different values of the relative cutoff (E_{cut}^{rel}). Red squares $E_{cut}^{rel} = 30$ Ry, green triangles $E_{cut}^{rel} = 40$ Ry and black circles $E_{cut}^{rel} = 50$ Ry. The non-monotonic behavior of the curves is the result of fortuitous cancellation of errors.

Lattice parameter optimization has been carried out for the NH_3 and CO_2 crystals. The employed procedure is approximate, as MP2 gradients and stresses are currently not available in CP2K, but is similar to the procedure in Ref. [107]. First, structures have been relaxed at a DFT/B3LYP level with the cc-TZVP basis set for various values of the lattice parameter. Next, $E_{coh}^{CP}(V)$ has been computed for each of these geometries. Finally, these results have been fitted with a third order Birch-Murnaghan equation of state in order to get the equilibrium cohesive energy and lattice parameter.

To assess the accuracy of computed cohesive energies, these values have been compared to the experimental sublimation enthalpies ($\Delta H(s)$). However, it has to be emphasized that this comparison includes theoretical bias and is subject to experimental error. Indeed, for non-volatile compounds, the sublimation enthalpies can be hard to measure, and can be subject to several kJ/mol error. Experimental sublimation enthalpies usually are obtained at high temperature, while the cohesive energy is a zero temperature property. Only in a few cases can experimental sublimation enthalpies be extrapolated down to 0K, CO_2 being such an example[114]: 26.8kJ/mol at 0K, 25.2kJ/mol at 195K. Even at 0K, zero point energy differences between gas and crystal should be taken into account, and in fact the anharmonicity of molecular crystals implies that also lattice parameters might need to be corrected for quantum effects to be truly comparable with experiment. Such corrections require the calculation of vibrational or phonon properties, see for example Ref. [115] for an early example based on a force field description of molecular crystals, and this has not been attempted here.



(a) **(b)**

Figure 3.6: Measured speed up and efficiency for the calculation of the GPW-MP2 energy. (a) Speed up measured with respect to 1600 processes for NH_3 crystal and with respect to 3200 processes for CO_2 crystal. (b) Efficiency measured with respect to 1600 processes for NH_3 crystal and with respect to 3200 processes for CO_2 crystal.

3.4.2 GPW-MP2 Accuracy

In order to judge the impact of the PW cutoff (E_{cut}), and the multi-grid relative cutoff (E_{cut}^{rel}) on the accuracy of the GPW-MP2 energy, calculations with various values for these parameters have been performed and summarized in Figure 3.5. The benchmark set is based on 10 molecules (from Figure 3.4 and H_2O) in the gas phase with the cc-TZVP basis, for which reference energies have been obtained using a traditional direct MP2 algorithm based on analytic four center integrals over atomic orbitals. The root-mean-square deviation (RMSD) of the absolute difference between the GPW-MP2 energy and the traditional MP2 implementation is used as a measure of the error. Tight values $\epsilon_{filter} = 10^{-12}$ and $\epsilon_{grid} = 10^{-12}$ together with a 17Å cubic cell and cluster boundary conditions have been used, to guarantee convergence with respect to these parameters. It can be observed that the GPW-MP2 energy converges rapidly with respect to both parameters E_{cut} and E_{cut}^{rel}, in particular if one realizes that the time for the integration in the MP2-GPW algorithm grows slowly as $E_{cut}^{\frac{3}{2}}$. Not unexpectedly, both parameters need to be increased simultaneously in order to obtain accurate results. The combinations 300/50Ry, 250/40Ry, and 200/30Ry yield errors of approximately 10^{-6}, 10^{-5}, and 10^{-4} Hartree respectively. Note again that E_{cut} depends on the largest exponent in the basis set used, while E_{cut}^{rel} is system independent.

3.4.3 Performance of the Parallel Algorithm

To assess the performance of the parallel algorithm, test calculations on molecular crystals of NH_3 and CO_2 have been performed. For NH_3 a cc-TZVP quality basis set has been chosen, while CO_2 has been described with a cc-QZVP basis. Each supercell contains

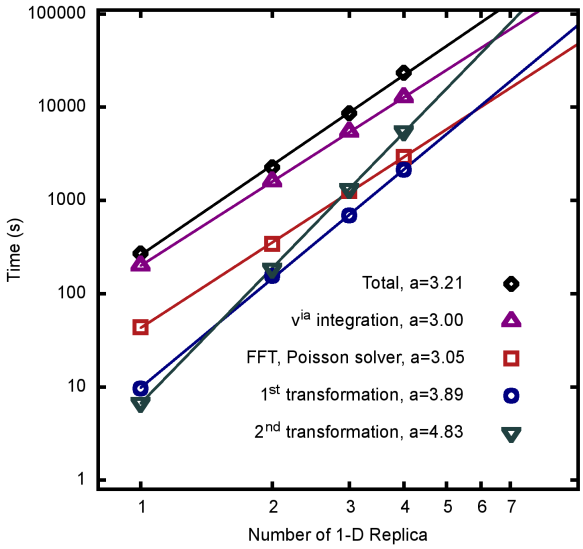


Figure 3.7: Shown is the time spent in the various significant parts of the GPW-MP2 energy calculation, as a function of the number of replicas of the supercell, containing 32, 64, 96 and 128 molecules of NH3 respectively. Lines represent a linear two-parameter fit of the form $y = bx^a$. The values of a for each task are reported in the legend.

32 molecules, resulting in 2272 and 5184 basis functions respectively The speed up and the parallel efficiency for the two test cases are reported in Figure 3.6. For both benchmarks the algorithm shows very good parallel scalability in a wide range. In particular in the case of CO_2, the efficiency remains higher than 80% even for the 102400 processes run. For the NH_3 calculation, the number of ia pairs (274560) becomes similar to the number of processes, making an even distribution of the pairs more difficult. Additionally, the overhead of initializing grids and matrices becomes non-negligible leading to a efficiency of 70%. At full scale-out, the MP2 energy calculation required a wall-time of 74 and 518 seconds, respectively.

3.4.4 System Size Scaling

In order to validate the expected timings for the important parts of the GPW-MP2 algorithm shown in Table 3.1, we have performed calculations on systems of increasing size. The test system is the supercell of NH_3, containing 32 molecules, with a cc-DZVP basis. This supercell has been replicated in one dimension, yielding an orthorhombic supercell containing up to 128 molecules. The timings, as obtained from runs on 1200 cores, are shown in Figure 3.7, together with a fit of the form $y = bx^a$. The measured exponents agree very well with the expected values as reported in Table 3.1, being very close to 3 for the integration and the Poisson solver, 4 for the first and 5 for the last index transformation. In addition to giving information about the scaling behavior, insight about the prefactor can be inferred from the graph. In particular, the last index transformation has a very small prefactor and an extrapolation suggest that it will only dominate for systems containing more than 200 molecules. Note that in an RI-MP2 calculation, this

Table 3.2: Hartree-Fock and MP2 contributions to the counterpoise corrected cohesive energies in mE_h of LiH at the experimental geometry for various basis sets and cell sizes. The text discusses how the extrapolated numbers (italic type) have been obtained.

	$2\times2\times2$	$3\times3\times3$	$4\times4\times4$	$5\times5\times5$	Extr. ($E_X^{n\to\infty}$)
			Hartree-Fock		
cc-DZVP	-139.90	-138.77	-138.37	-138.36	
cc-TZVP	-132.84	-132.20	-132.03	-132.05	
cc-QZVP	-132.60	-132.02			
			MP2 all-electron		
cc-DZVP	-29.26	-30.31	-30.58	-30.66	*-30.76*
cc-TZVP	-38.00	-39.34	-39.68		*-39.91*
cc-QZVP	-40.55	-41.98	*-42.34*		*-42.59*
Extr. ($E_{X\to\infty}^{n\to\infty}$)					*-44.09*
			MP2 valence-only		
cc-DZVP	-27.89	-28.83	-29.06	-29.13	*-29.22*
cc-TZVP	-36.23	-37.42	-37.73		*-37.94*
cc-QZVP	-38.68	-39.96	*-40.28*		*-40.51*
Extr. ($E_{X\to\infty}^{n\to\infty}$)					*-41.93*

term has a larger prefactor, approximately given by the ratio of basis function in the auxiliary and primary basis. For all system sizes tested, the cost is currently dominated by the integration routine, yielding an apparent overall scaling exponent of the GPW-MP2 algorithm of 3.21.

3.4.5 Solid LiH

The LiH crystal has become a well studied benchmark system for condensed phase electronic structure calculations. It has favorable characteristics such as a large bandgap, a simple unit cell, and the absence of heavy elements. This system is thus useful to judge the quality of a method or implementation before more challening systems, for example with large unit cells, small band gaps, or complicated chemistry are attempted. Various results for the MP2 contribution to the cohesive energy of LiH have recently been presented in literature [50, 116–119]. In this section, the computation of the MP2 cohesive energy is used to validate the GPW-MP2 method and to analyze system size and basis set convergence behavior.

The results obtained for the Hartree-Fock and MP2 contributions to the counterpoise corrected cohesive energy of LiH are summarized in Table 3.2. Calculations have been performed up to the $5\times5\times5$ cell for the cc-DZVP basis, $4\times4\times4$ for cc-TZVP and $3\times3\times3$ for cc-QZVP. MP2 results have been computed for valence-only and all-electron correlation. The largest cells, for each basis, contain roughly 10000 basis functions, and the $5\times5\times5$ cell contains 1000 atoms. The Hartree-Fock results agree well with previously reported pseudo-potential-free calculations performed with CP2K[106] ($-131.95\ mE_h$). The MP2

Table 3.3: Ratio between the MP2 contributions to the cohesive energy of LiH for two consecutive cell sizes $\left(\frac{E_X^n}{E_X^{(n+1)}}\right)$, with X=D,T,Q.

	MP2 all-electron			MP2 valence-only		
	$\frac{E_X^2}{E_X^3}$	$\frac{E_X^3}{E_X^4}$	$\frac{E_X^4}{E_X^5}$	$\frac{E_X^2}{E_X^3}$	$\frac{E_X^3}{E_X^4}$	$\frac{E_X^4}{E_X^5}$
cc-DZVP	0.9654	0.9911	0.9974	0.9674	0.9921	0.9975
cc-TZVP	0.9661	0.9914		0.9680	0.9920	
cc-QZVP	0.9660			0.9679		

results depend rather strongly on the basis set and cell size used, and extrapolation is thus necessary. In the following discussion, the MP2 contribution is labeled E_X^n, with X denoting the basis (D,T,Q,∞) and n the number of repeated unit cells (2,3,4,5,∞).

As shown in Table 3.3, the ratio between the cohesive energies for two consecutive cell sizes $\left(\frac{E_X^n}{E_X^{(n+1)}}\right)$ converges quickly with basis set size. This is equivalent to saying that the system size effect is similar for various basis sets. For example, the difference of $\frac{E_X^2}{E_X^3}$ in going from the cc-TZVP to the cc-QZVP basis is around 10^{-4} for both the all-electron and valence-only case. Based on this observation, E_Q^4 is obtained as the extrapolated result $E_Q^3 \times \frac{E_T^4}{E_T^3}$. Note that the difference between the computed result E_Q^3 and the extrapolated $E_Q^2 \times \frac{E_T^3}{E_T^2}$ is approximately $4\mu E_h$, validating this approach. In order to estimate the size converged limit, an extrapolation for the cell size going to infinity ($E_X^{n\to\infty}$), for a given basis set, has been performed by a linear fit of the equation $E_X^n = E_X^{n\to\infty} + S(n \times a)^{-3}$ where n is the number of repeated unit cells, and a the lattice constant. The choice of exponent (-3) is justified by the fact that the long range behavior of the MP2 pair energy follows the London law C_6/d_{ij}^6, with d_{ij} being the distance between the center of two distributions[59], and by integrating over all pairs in the crystal for which $d_{ij} \geq d$. Finally, the size extrapolated results for each of the basis sets are extrapolated to the basis set limit. This basis set extrapolation is based on the cubic interpolation formula[120] $E_X^{n\to\infty} = E_{X\to\infty}^{n\to\infty} + AX^{-3}$ (X=2,3,4 for cc-DZVP, cc-TZVP and cc-QZVP respectively). The fits are shown in Figure 3.8 and and the extrapolated values reported in Table 3.2. We note that the extrapolation for the cell size going to infinity yields an asymptotic standard error of the order of few μE_h, but for the basis set extrapolation the estimated error is on the order of a tenth of mE_h. Finally, the MP2 contribution to the cohesive energy of LiH is estimated to be -44.09 mE_h in the all-electron case and -41.93 mE_h for valence-only correlation.

As summarized in Table 3.4, the estimates presented here are in good agreement with other values reported in literature. For the all-electron case, excellent agreement with the result reported by Marsman and co-workers [50] is observed, the deviation is only 0.1 mE_h.

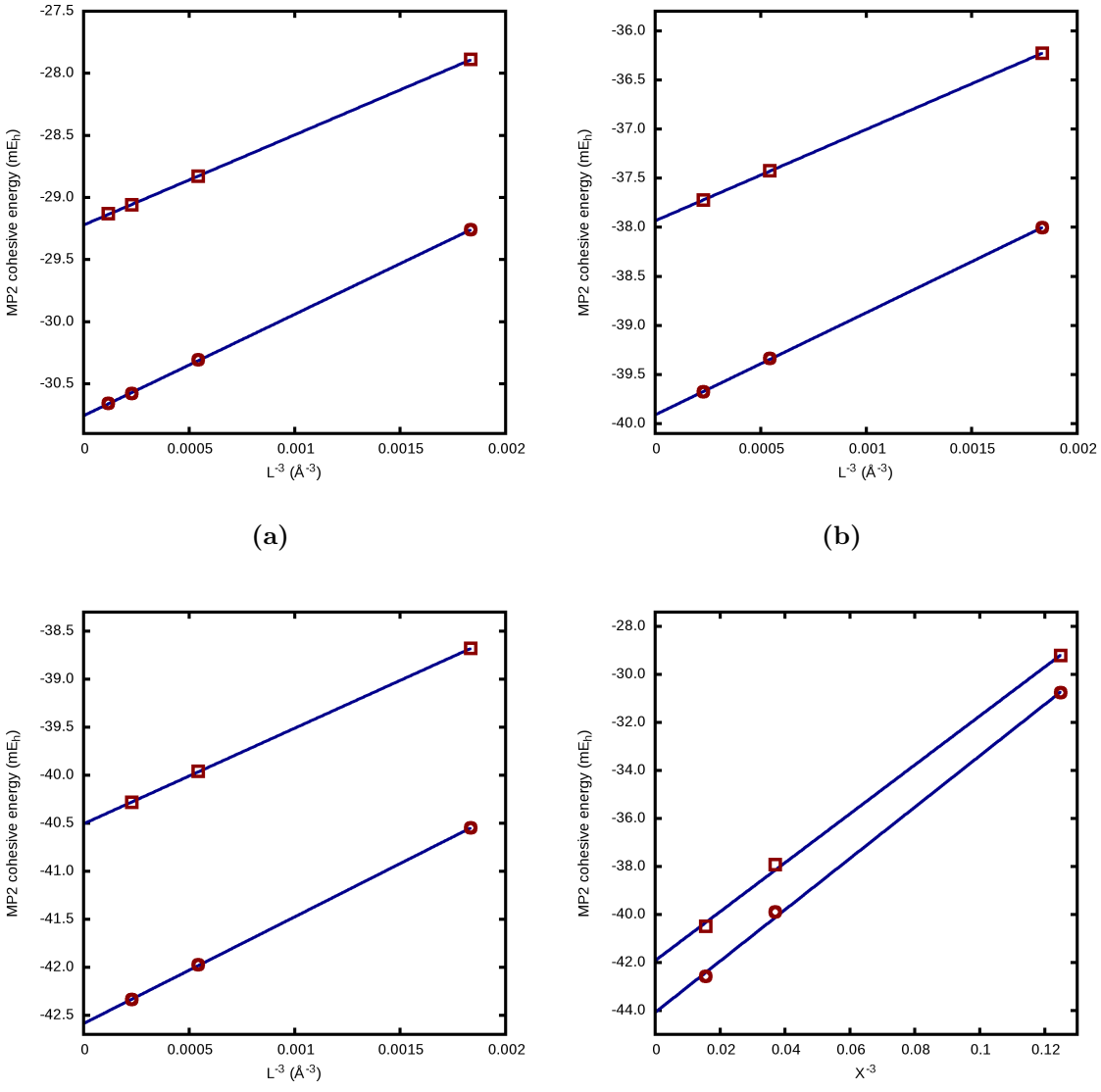


Figure 3.8: Convergence of the all-electron (circles) and valence-only (squares) MP2 contribution to the cohesive energy of LiH with respect to the cell size and basis set. L is the length of the cell edge ($L = n \times a$) and $X = 2, 3$ and 4 for cc-DZVP, cc-TZVP and cc-QZVP respectively. (a) Extrapolation with respect to the cell size for the cc-DZVP basis set. (b) Extrapolation with respect to the cell size for the cc-TZVP basis set. (c) Extrapolation with respect to the cell size for the cc-QZVP basis set. (d) Extrapolation with respect to the basis set for the size-converged MP2 cohesive energy.

Table 3.4: Comparison of the basis set limit estimation of the MP2 contribution to the cohesive energy of LiH at the experimental geometry (a = 4.084 Å). All energies in mE_h. Results calculated with different approaches, reported by different authors.

Code/Method	E_{coh}^{MP2}	
Hierarchical Method [a]	all-electron	-43.43
VASP [b]	all-electron	-43.99
CP2K [c]	all-electron	-44.09
CRYSCOR [d]	valence-only	-41.34
Incremental Scheme [e]	valence-only	-41.55
CP2K [c]	valence-only	-41.93

[a] Nolan and co-workers [118]
[b] Marsman and co-workers [50]
[c] This work
[d] Usvyat and co-workers [117]
[e] Stoll and co-workers [116]

The other values show larger discrepancies, up to 0.6 millihartree. We believe that these differences are due to the delicate nature of the basis set extrapolation. In the condensed phase, a robust extrapolation procedure might be easier to establish for a plane waves basis, and a detailed analysis for such an approach has recently been presented in Ref. [119]. Without extrapolation and for very similar basis sets, better agreement is observed. For example, the value of the valence-only MP2 energy reported by Stoll and co-workers[116] at the triple-zeta level is -38.02 mE_h, very close to the -37.94 mE_h obtained with the cc-TZVP basis. With the "basis set A" proposed by Usvyat and co-workers [117] (with small adaptions of the core-valence part of the basis to the pseudopotentials), we find -38.97 mE_h for the valence-only contribution, in good agreement with -39.08 mE_h reported in ref [117].

3.4.6 NH_3 and CO_2 Molecular Crystals

The molecular crystals of NH_3 and CO_2 present two useful benchmark systems, as they differ in the nature of their interaction: NH_3 is hydrogen bonded, while CO_2 is not. Furthermore, results for these systems can be compared to results presented in Ref. [107] and obtained by LMP2 as implemented in CRYSCOR. The effect of the basis set and the supercell size on the cohesive energy at the experimental geometry are reported in Table 3.5. Not unexpectedly, it can be seen that cc-DZVP yields poor results for the cohesive energy, and a cc-TZVP basis is required to yield a cohesive energy within approx. 5 kJ/mol of the basis set extrapolated result. However, the smaller cc-DZVP basis allows for larger systems, and hence can be used to study the size dependence of the result. Indeed, since our results are obtained at the Γ-point only, the size of the supercell matters. Fortunately, we see that the difference between the cohesive energies obtained 2x2x2 and 3x3x3 unit cells (32 and 108 molecules respectively) is rather small, a few kJ/mol, smaller than the difference between cc-TZVP and basis set extrapolated results.

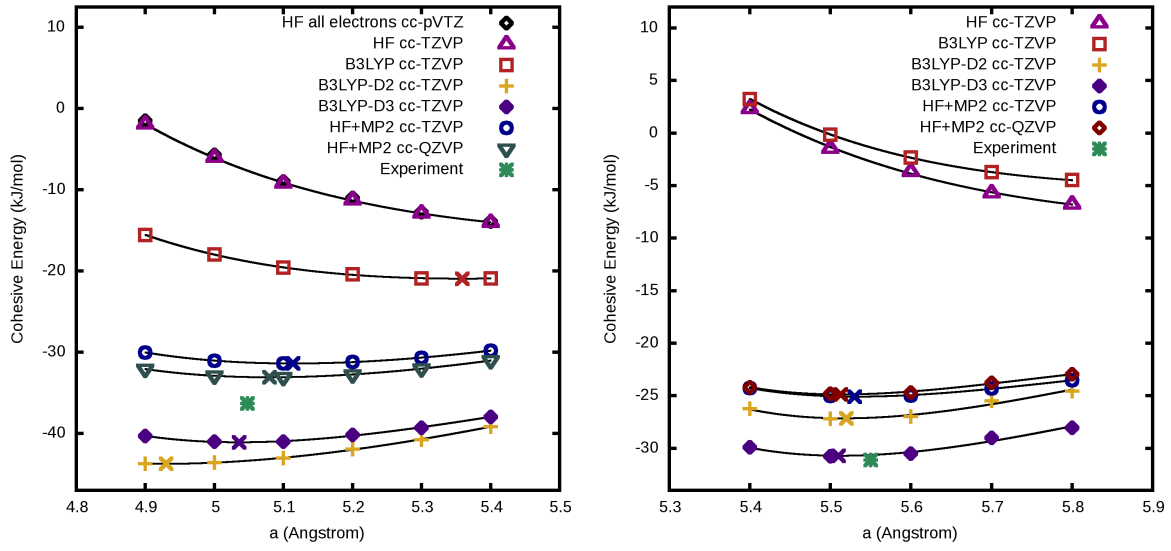
Table 3.5: Counterpoise corrected cohesive energy (E_{coh}^{CP}) in kJ/mol for the NH_3 and CO_2 crystals. Results have been calculated employing different basis sets and system sizes at the experimental geometry. Values in parenthesis refer to the HF+LMP2 results from Ref. [107] as obtained with the CRYSCOR program, using basis sets of comparable quality.

	NH_3		CO_2	
	HF	HF+MP2	HF	HF+MP2
cc-DZVP	-8.43	-25.78 (-22.6)	-7.48	-8.92 (-8.9)
cc-DZVP $(3 \times 3 \times 3)$ [a]	-8.80	-27.89	-7.47	-11.46
cc-TZVP	-5.98	-30.93 (-29.6)	-5.86	-20.95 (-19.3)
cc-QZVP	-5.60	-32.76 (-32.3)	-5.91	-23.28 (-24.4)
Extrapolated [b]	-5.52	-33.93	-5.99	-26.09
Experiment [c]		-36.3		-31.1

[a] Calculation performed with supercell $3 \times 3 \times 3$ instead of $2 \times 2 \times 2$.

[b] (T-Q) extrapolation toward the basis set limit [120].

[c] Values from Ref. [107], see also Ref. [121] for NH_3 and Ref. [122, 123] for CO_2.



(a) **(b)**

Figure 3.9: Lattice parameter optimization curves for NH_3 (a) and CO_2 (b), computed at different level of theory and different basis sets. The crosses represent the location of the minimum point for each curve.

Table 3.6: Equilibrium counterpoise corrected cohesive energy (E_{coh}^{CP} in kJ/mol per molecule) and lattice parameter (a in Å) for the NH_3 and CO_2 crystals calculated employing different methods. The basis set is cc-TZVP, except when specified otherwise.

	NH_3		CO_2	
	a	E_{coh}	a	E_{coh}
HF	5.98	-15.7	6.05	-5.00
B3LYP	5.36	-20.9	6.20	-8.20
B3LYP-D2	4.93	-43.7	5.52	-27.2
B3LYP-D3	5.04	-41.1	5.51	-30.7
HF+MP2	5.11	-31.4	5.53	-25.1
HF+MP2 [a]	5.08	-33.1	5.51	-24.9
HF+LMP2 [b]	5.02	-36.6	5.59	-26.6
Exp.	5.048	-36.3	5.55	-31.1

[a] Calculated with cc-QZVP basis set.

[b] Reference [107], calculated with aug(d,f)-TZPP basis set.

Note that the HF results are less sensitive to both system size and basis set effects. Of course, the Hartree-Fock level of theory is a poor description for these systems, since the MP2 correlation contributes almost 80% of the cohesive energy. Careful estimates of the MP2 cohesive energy could combine the size effect as computed with smaller basis sets with basis set extrapolated results for smaller supercells, as we have illustrated for LiH. Furthermore, our results are in good agreement with those reported in Ref.[107], despite the various differences in methodology, such as the use of pseudopotentials, corresponding basis sets, local MP2, etc. Lattice parameters are within .1Å of the HF+LMP2 results and of experiment.

In Figure 3.9, lattice parameter optimization curves are reported as obtained at various levels of theory. In the case of NH_3, the quality of the pseudopotential approximation has been verified as illustrated by the excellent agreement between the HF results as obtained with pseudopotentials and the corresponding cc-TVZP basis, and an all-electron calculation employing the Dunning cc-pVTZ basis. For both NH_3 and CO_2, the DFT/B3LYP and HF results predict an equilibrium lattice parameter much larger than the experimental one, and a cohesive energy that is far from the experimental one. The HF results are significantly improved by the MP2 correction, and the same can be said for the dispersion corrections in the case of B3LYP. The calculated equilibrium values of a and E_{coh} are summarized in Table 3.6. The MP2 lattice constants are within approx. 1% of the experimental results, and while the same holds for B3LYP-D3 [124], a larger error is observed with B3LYP-D2 [125]. The cc-TZVP and cc-QZVP MP2 results are very similar, suggesting that the former might be a cost effective choice for these calculations.

3.4.7 Cohesive Energy of Molecular Crystals from MP2 and double hybrid DFT

Finally, cohesive energies for the remaining molecular crystals reported in Figure 3.4 are presented in Table 3.7, in all cases using the experimental crystal cell, and employing the cc-TZVP basis. Here, a wider set of theoretical methods is employed, including MP2 as well as double hybrid functionals. These values are compared to the experimentally measured sublimation enthalpies, with the caveats mentioned above. The RMSD, measured over this relatively small test set, is reported as well. First, HF and DFT/B3LYP methods display the largest RMSD, demonstrating that these methods are not suitable for computing, even to a qualitative level, the lattice energy of molecular crystals. Small molecules that interact mostly via hydrogen-bonds and dipole-dipole electrostatics are described best, but van der Waals dominated complexes can have even a negative computed lattice energy. On the other hand, the D2 and D3 dispersion corrections appreciably improve the bare B3LYP results, for all cases the computed lattice energy goes closer to the experimental value, decreasing significantly the RMSD. However, both B3LYP-D2 and B3LYP-D3 tend to overbind the crystals, B3LYP-D3 being slightly worse. With an RMSD of 16.9 kJ/mol, the lattice energies calculated at the MP2 level outperform the DFT results. The molecules that display the largest deviations, Benzene and Pyromellitic Dianhydride, have a large π electronic delocalization, which is known to be unfavorable for the performance of MP2. Relaxation of the gas phase geometry has a large effect on the cohesive energies for those systems with strong interactions in the crystal, and is only negligible for benzene. Also in this case, there is good agreement between the computed GPW-MP2 energies (non relaxed case) and the LMP2 calculations reported in Ref. [111].

The spin component scaled variants of the MP2 method tend to improve slightly over the performance of standard MP2, the MP2-SCS method of Grimme[126] yielding the smallest RMSD. This is largely due to the better agreement observed for Pyromellitic Dianhydride, for which the discrepancy is reduced from ∼40 to ∼20 kJ/mol. Interestingly, the MP2-SCS(MI) of Distasio *et al.*[128], which has been parameterized explicitly for molecular interactions using the S22 database [130] as training set, does not perform best for the molecular interactions in these crystals. It is the best method for benzene, which is also present in the S22 database in a similar configuration as the one found in the crystal, but it is less accurate for formic acid, which is present in S22 in a fairly different geometry. This suggests that a database of accurate cohesive energies for molecular crystals would complement the S22 set by providing a wider range of molecular geometries, and thus provide valuable input for the development of improved methods.

Two different double hybrid functionals have been tested: B2PLYP and DSD-BLYP with and without dispersion corrections [19]. The main difference between these two functionals is that the MP2-like term is scaled with a single parameter in B2PLYP, while independent parameters are used for the same spin and opposite spin MP2-like terms in DSD-BLYP, as is done in MP2-SCS. B2PLYP without dispersion correction shows a large RMSD. B2PLYP-D3, with an added D3 correction, improves slightly over MP2. Finally, the DSD-BLYP functional yields the best results for Pyromellitic Dianhydride, which is the source of large errors for other methods. Surprisingly, adding the D3 correction to DSD-BLYP actually increases the RMSD, increasing the errors for some compounds significantly.

Table 3.7: Counterpoise corrected cohesive energy ($-E_{coh}^{CP}$) in kJ/mol for the the molecular crystals of B = Benzene (fig Figure 3.4f), FA = Formic Acid (fig Figure 3.4c), SA = Succinic Anhydride (fig Figure 3.4e), D = 2,3-Diazanaphthalene (fig Figure 3.4g), PD = Pyromellitic Dianhydride (fig Figure 3.4h), U = Urea (fig Figure 3.4d), CT = Cyclotrimethylene-Trinitramine (fig Figure 3.4i). The sign of the E_{coh}^{CP} has been changed in order to be compared with the experimental sublimation enthalpies $\Delta H(s)$. (C) means that the gas phase geometry has not been relaxed but was constrained to the crystal geometry.

	B	FA	SA	D	PD	U	CT	RMSD
Exp $\Delta H(s)$ [a]	45	68	81	83	83	92	112	
LMP2 (C) [b]	56.6	63.2	87.0			108.6		
GPW-MP2 (C)	58.7	64.9	84.9	93.3	127.3	106.6	126.6	19.7
GPW-MP2	58.8	55.5	81.2	79.7	123.4	94.6	113.7	16.9
HF	-21.2	26.3	38.6	-5.7	31.3	55.8	49.8	58.1
B3LYP	-12.5	34.1	27.5	-7.0	15.8	64.2	17.3	65.1
B3LYP-D2	56.9	69.1	91.8	84.0	122.0	110.5	128.4	18.5
B3LYP-D3	60.5	71.9	94.7	87.3	128.1	111.6	131.0	21.5
B2PLYP [c]	14.7	44.8	51.0	74.2	60.0	78.4	58.8	29.2
B2PLYP-D3 [c]	53.4	64.7	86.4	77.3	119.8	103.7	119.9	15.6
DSD-BLYP [c]	31.6	54.4	67.6	50.6	88.8	90.9	90.8	17.2
DSD-BLYP-D3 [c]	56.9	67.4	90.8	82.8	128.1	107.4	130.9	20.2
MP2-SCS [d]	38.6	52.2	75.6	62.3	106.4	89.2	112.0	13.6
MP2-SOS [e]	28.5	50.6	72.9	53.6	97.9	86.7	111.2	15.8
MP2-SCS(MI) [f]	47.9	47.3	70.1	63.9	104.6	84.8	90.2	16.5
MP2-SCSN [g]	52.3	44.9	67.1	64.3	103.4	82.3	79.3	19.7

[a] Taken from the supporting information of Ref [111], see also `http://webbook.nist.gov/chemistry/`.

[b] Local MP2 calculation performed with the CRYSCOR program reported from Ref [111].

[c] The parameter for these double hybrid functional and their relative D3 corrections have been taken from Ref [19].

[d] Spin Component Scaled (SCS)[126] ($p_S = 1.2$, $p_T = 0.333$).

[e] Scaled Opposite Spin (SOS)[127] ($p_S = 1.3$, $p_T = 0$).

[f] Spin Component Scaled (Molecular Interaction) (SCS(MI))[128] ($p_S = 0.4$, $p_T = 1.29$).

[g] Spin Component Scaled for Nucleobases (SCSN)[129] ($p_S = 0$, $p_T = 1.76$).

3.5 Conclusions

In this chapter, a novel method for the calculation of the canonical MP2 energy of finite and extended systems has been presented. The crucial aspect of the method is that half-transformed electron repulsion integrals (ERIs) $(ia|\lambda\sigma)$ are directly calculated. This is possible using a mixed Gaussian and Plane Wave approach, which allows for computing the electrostatic potential v^{ia} of the occupied-virtual pair density ρ^{ia} in an auxiliary basis, and numerically integrating over products of basis functions $\lambda\sigma$. The method is naturally suited and robust for periodic systems and the numerical accuracy can be easily controlled. The corresponding algorithm shows excellent parallel performance up to 100000 processes, and allows for MP2 calculations of systems containing hundreds of atoms and thousands of basis functions in minutes. Benchmark calculations on solid LiH and molecular crystals have been performed to validate the GPW-MP2 method, and good agreement with literature results and, for most benchmarks, with experiment is obtained. These calculations also suggest that a database with reliable reference cohesive energies for molecular crystals could complement existing gas phase databases, and contribute to the development of improved methods and functionals for weak interactions. We believe that the GPW-MP2 method can now be used to study condensed phase systems with a few hundred atoms per unit cell, including not only crystals but also systems without symmetry such as molecules on surfaces and liquids. Advanced techniques, such as RI-MP2 and local MP2, and adaptation to new hardware, such as accelerators, are likely to further improve upon the method presented here.

Chapter 4

Electron correlation in the condensed phase from a resolution of identity approach based on the Gaussian and Plane Waves scheme [131]

The second-order Møller-Plesset perturbation energy (MP2) and the Random Phase Approximation (RPA) correlation energy, are increasingly popular post-Kohn-Sham correlation methods. In this chapter, a novel algorithm based on a hybrid Gaussian and Plane Waves (GPW) approach with the resolution-of-identity (RI) approximation is presented for MP2, scaled opposite-spin MP2 (SOS-MP2) and direct-RPA (dRPA) correlation energies of finite and extended system. The key feature of the method is that the three center electron repulsion integrals $(\mu\nu|P)$ necessary for the RI approximation are computed by direct integration between the products of Gaussian basis functions $\mu\nu$ and the electrostatic potential arising from the RI fitting densities P. The electrostatic potential is obtained in a plane waves basis set after solving the Poisson equation in Fourier space. This scheme is highly efficient for condensed phase systems and offers a particularly easy way for parallel implementation. The RI approximation allows to speed up the MP2 energy calculations by a factor 10 to 15 compared to the canonical implementation, but still requires $O(N^5)$ operations. On the other hand, the combination of RI with a Laplace approach in SOS-MP2 and an imaginary frequency integration in dRPA reduces the computational effort to $O(N^4)$ in both cases. In addition to that, our implementations have low memory requirements and display excellent parallel scalability up to ten thousands of processes. Furthermore, exploiting graphics processing units (GPU), a further speed-up by a factor ~ 2 is observed compared to the standard only CPU implementations. In this way, RI-MP2, RI-SOS-MP2 and RI-dRPA calculations for condensed phase systems containing hundreds of atoms and thousands of basis functions can be performed within minutes employing a few hundred hybrid nodes. In order to validate the presented methods, various molecular crystals have been employed as benchmark systems to assess the performance, while solid LiH has been used to study the convergence with respect to the basis set and system size in the case of RI-MP2 and

RI-dRPA.

4.1 Introduction

Density Functional Theory (DFT) has become the most widely used quantum mechanical tool in chemistry and physics for predicting properties of materials ranging from molecules to condensed phase systems. The strength of DFT lies in a good compromise between accuracy and computational effort. In fact, existing implementations of DFT in its local and semi-local formulation display a computational complexity growing only $O(N) - O(N^3)$, where N represents the system size. However, going beyond these simple formulations is required to make significant progress on the way to high accuracy.

In this respect, Perdew *et al.* [3] have proposed a systematic classification of the existing approximate density functionals based on the information they employ. This classification is known as "Jacob's ladder". The fifth rung of the "Jacob's ladder" includes methods that make use of information from the occupied and virtual orbitals, and allows for taking into account the non-local dynamical electron correlation contributions, including the long-range van der Waals (vdW) dispersion interactions. Most of the 5^{th} rung approaches are based on either the random phase approximation (RPA), introduced in the framework of DFT via the so-called adiabatic-connection fluctuation-dissipation theorem (ACFD) [4–12], or many body perturbation theory (MBPT) in the form of second-order Møller-Plesset (MP2) [13–19] perturbation theory.

The drawback connected with these improved functionals is the computational cost, that grows significantly compared to standard DFT approaches of lower rung. In fact, even if low-order scaling methods exist [48, 49, 53–61, 65–67, 132], the formal computational effort scales between $O(N^4)$ to $O(N^5)$ with respect to systems size. Furthermore, these methods exhibit the same slow energy convergence as correlated wave-function methods due to the electron coalescence cusp [120, 133, 134], implying that these calculations need larger basis sets than standard DFT to reach a similar convergence, unless explicitly correlated [74] or range separation [135, 136] methods are employed. These are the reasons that have impeded the widespread use of fifth rung density functional approaches, limiting their application to small size systems.

In order to perform these calculations with acceptable time to solution and to extend the applicability of these methods to larger systems, the development of efficient algorithms as well as computer implementations, suitable for massively parallel machines, are of prime interest. In this respect, the resolution of identity approximation (RI) has shown to be a powerful technique to reduce the prefactor and the scaling for both RPA and MP2 calculations [43, 46, 47, 68–73, 127, 137, 138], furthermore several parallel MP2 and RI-MP2 implementations have been proposed [75–85, 139] displaying good scalability up to few hundred cores, but none of them have been yet reported in the case of RPA.

Here, we present a novel algorithm for the resolution of identity approximation based on Gaussian and Plane Wave (GPW) approach [44]. Of central importance in the RI approximation is the computation of the three center electron repulsion integrals of the type $(\mu\nu|P) = \sum_R (\mu\nu|R) L^{-1}_{PR}$, where $\mathbf{L}$ is obtained by the Cholesky decomposition of the two center electron repulsion integrals $(Q|R)$ over auxiliary Gaussian basis functions. The key aspect of our method is that the three center electron repulsion integrals of the

type $(\mu\nu|P)$ are computed by direct integration between the product of the Gaussian basis function $(\mu\nu|$ and the electrostatic potential arising from the RI fitting densities $(P| = \sum_R (R|L^{-1}_{PR}$. Within the GPW approach, the electrostatic potential is obtained by solving the Poisson equation in Fourier space, after expanding the the RI fitting densities in a plane wave (PW) auxiliary basis set. Note that the PW auxiliary basis is a natural choice for periodic systems, but it can equally be used for gas phase or surface calculations. Indeed, once the density is specified on a regular grid efficient methods are available for solving the Poisson equation with free (for example cluster or slab) boundary conditions[93–95]. The GPW scheme is highly efficient, especially for condensed phase systems, and particularly suitable for parallel implementation, since the $(\mu\nu|P)$ integrals can be computed in a communication free way simply by distributing independent $(P|$ vectors over processes.

Alternative RI methods suitable for the condensed phase that do not rely on a PW auxiliary basis have been described in literature. A first approach, density fitting (DF), has been reported for the calculation of the local-MP2 energy for periodic systems [59, 72, 73, 140]. In this approach, the large majority of the fitting functions (FFs) are chosen to be Poisson-type functions (PTFs), simplifying the calculation of the Coulomb two-electron integrals into overlap integrals and allowing for rapidly converging lattice sums [72]. Nevertheless, PTF need to be augmented with a small set of Gaussian type functions to be able to describe multipole moments. A second approach, which employs an RI expansion for calculating the Coulomb term in periodic systems, has been proposed by Burow and coworkers[141]. Here, the lattice sums are partitioned into a near-field and far-field part, the first treated analytically, the latter using the periodic continuous fast multipole method[142, 143]. Compared to the GPW scheme, these methods are more involved. However, the simplicity of the GPW method has as a drawback that all-electron calculations are not possible, and that pseudopotentials have to be employed in order to have densities that are smooth. The Gaussian and Augmented Plane Wave (GAPW) scheme [96, 97] overcomes this limitation and is suitable for all-electron calculations, but our RI implementation is currently limited to the GPW method only.

We apply the RI approximation to the computation of the MP2, scaled opposite-spin MP2 (SOS-MP2) [126, 127] and direct-RPA (dRPA) [34, 36, 43] correlation energies, for condensed and finite systems. In the case of the computation of the MP2 energy the RI approximation reduces drastically the prefactor for the overall calculation, giving speed-ups, compared to standard implementations, up to a factor fifteen, but still retaining the formal $O(N^5)$ scaling. On the other hand, in the SOS-MP2 and dRPA cases, the RI approximation not only speeds-up the calculation of the integrals, but also allows, in both cases, to reduce the scaling to $O(N^4)$. This is achieved in the SOS-MP2 case by a Laplace approach [62, 63, 127] and in dRPA by reformulating the correlation energy expression in terms of an imaginary frequency integral [43].

The parallel algorithms for the calculation of the RI integrals within the GPW approach and for the computation of the MP2, SOS-MP2 and dRPA correlation energies within the RI approximation have been implemented in CP2K[90]. The presented parallel implementations display excellent parallel scalability and efficiency up to several thousands of processes and allow for correlation energy calculation on systems comprised of hundreds of atoms and thousands of basis functions within minutes. We also report the performance of hybrid implementations making use of graphics processing units (GPUs) showing further

speed-up compared to a standard CPU-only implementation.

The remainder of the chapter is organized as follow: first the basic theoretical framework of each method together with the employed notation are introduced, then the the parallel implementations and performance are discussed in detail. Finally, several benchmark calculations for different condensed phase systems are reported in order to validate the proposed approaches.

4.2 Theory

In this section the theoretical framework of the methods is briefly presented referring to the original works for more details. First the resolution of the identity approximation for two electron repulsion integrals (ERIs) is introduced and then its application to the different correlation methods is formulated giving the working expressions employed in the implementations. Finally, the Gaussian and Plane-Waves scheme is reviewed and its application together with the RI approximation is discussed. The following index notation has been adopted: $i, j, k, \ldots$ refer to canonical occupied molecular orbitals (MOs), $a, b, c, \ldots$ to canonical virtual MOs, $\mu, \nu, \lambda, \ldots$ to atomic orbital basis set functions (AO), $P, Q, R, \ldots$ to auxiliary basis set functions (AUX). The one electron MO and AO functions are symbolized respectively with ψ and ϕ. The number of occupied and virtual orbitals is denoted by o and v, while the total number of primary and auxiliary basis functions as n and N_a. In order to express, in general, the system size, the symbol N is used.

4.2.1 The Resolution of the Identity Approximation

The two electron repulsion integrals, in Mulliken notation, of the type $(ia|jb)$ are of central importance for all the methods presented in this chapter. Within the RI approximation [144, 145], based on the Coulomb metric [146], these integrals are factorized according to:

$$(ia|jb)_{RI} = \sum_{PQ} (ia|P)(P|Q)^{-1}(Q|jb) \tag{4.1}$$

here the $(P|Q)^{-1}$ are the matrix elements of the inverse of the matrix $(P|Q)$, where each element is given by

$$(P|Q) = \int\int \phi_P(\vec{r}_1) \frac{1}{|\vec{r}_1 - \vec{r}_2|} \phi_Q(\vec{r}_2) d\vec{r}_1 d\vec{r}_2. \tag{4.2}$$

The auxiliary basis set size N_a grows only linearly with the system size [69].

The main advantage of the RI approximation is that four center integrals of the type $(ia|jb)$ are computed from three and two center ERIs. This allows to strongly reduce the effort for the integral computation without significant loss of accuracy [69, 147].

Since the $(P|Q)$ matrix is positive definite the calculation of $(P|Q)^{-1}$ can be efficiently performed by a Cholesky decomposition of $(P|Q)$

$$(P|Q) = \sum_{R} L_{PR} L^T_{RQ} \tag{4.3}$$

followed by the efficient inversion of the triangular matrix $\mathbf{L}$, such that:

$$(P|Q)^{-1} = \sum_R L_{PR}^{-T} L_{RQ}^{-1}. \tag{4.4}$$

In this way the factorization of the $(ia|jb)$ integrals can be expressed in a compact form as:

$$(ia|jb)_{RI} = \sum_P B_P^{ia} B_P^{jb}. \tag{4.5}$$

Here $\mathbf{B}$ is a matrix with ov rows and N_a columns, given by:

$$B_P^{ia} = \sum_R (ia|R) L_{PR}^{-1}. \tag{4.6}$$

Since the three center integrals $(ia|R)$ are computed starting from integrals over AOs $(\mu\nu|R)$, the final expression for the B_P^{ia} elements reads:

$$(ia|P) = \sum_\nu C_{\nu a} \sum_\mu C_{\mu i} \sum_R (\mu\nu|R) L_{PR}^{-1} \tag{4.7}$$

where $\mathbf{C}$ is the MO coefficient matrix.

The computation of the $\mathbf{B}$ matrix can thus be summarized as follows. First the two center integrals $(P|Q)$ are computed and from that, via Cholesky decomposition and triangular inversion, $\mathbf{L^{-1}}$. These two steps formally scale $O(N^2)$ and $O(N^3)$, respectively. Then the three center integrals $(\mu\nu|R)$ are computed and subsequently transformed using the $\mathbf{C}$ and $\mathbf{L^{-1}}$ matrices (Equation (4.7)). In this case the integral computation requires formally $O(N^3)$ operations while the integral transformations scale $O(N^4)$.

This means that, within RI approximation, the asymptotically dominating time determining step in computing $\mathbf{B}$ is the index transformation scaling formally $O(N^4)$.

4.2.2 RI-MP2 Method

In Second Order Møller-Plesset Perturbation Theory, the correlation energy $E^{(2)}$ for a closed shell system is given by:

$$E^{(2)} = -\sum_{i \le j}^{o} (2 - \delta_{ij}) \sum_{ab}^{v} \frac{(ia|jb)[2(ia|jb) - (ib|ja)]}{\epsilon_a + \epsilon_b - \epsilon_i - \epsilon_j}. \tag{4.8}$$

where ϵ_a and ϵ_i are orbital energies.

In a canonical MP2 energy algorithm the time limiting step is the computation of the $(ia|jb)$ integrals obtained from the ERIs over AO $(\mu\nu|\lambda\sigma)$ via four consecutive integral transformations:

$$(ia|jb) = \sum_\mu C_{\mu i} \sum_\nu C_{\nu a} \sum_\lambda C_{\lambda j} \sum_\sigma C_{\sigma b} (\mu\nu|\lambda\sigma). \tag{4.9}$$

The computational effort for each of the four quarter transformations, if the occupied orbitals are transformed before the virtual, and sparsity is not considered, is $O(on^4)$,

$O(o^2n^3)$, $O(o^2vn^2)$ and $O(o^2v^2n)$, making the MP2 energy calculation a method scaling as $O(N^5)$ [148].

The application of the RI approximation to the MP2 energy calculation is straightforward [68]. It consists simply in the replacement of the $(ia|jb)$ integrals with the approximated $(ia|jb)_{RI}$ given in equation Equation (4.5). The computation of the $(ia|jb)_{RI}$ requires $O(o^2v^2N_a)$ operations implying that the RI-MP2 method is also scaling $O(N^5)$.

The main reason for the speed-up observed in RI-MP2 lies in the fact that ERIs over four indices are replaced by ERIs over two and three indices, which strongly reduces the effort in the integral computation part. However, this means that the speed-up becomes less and less pronounced increasing the system size. Indeed, for large systems, when the $O(N^5)$ step dominates, RI-MP2 and standard MP2 have a similar computational cost.

4.2.3 Laplace Transform Scaled Opposite-Spin RI-MP2 Method

The scaled opposite-spin second order correlation energy $E^{(2)}_{SOS}$ is defined as [126, 127]:

$$E^{(2)}_{SOS} = c_{SOS} E^{(2)}_{OS} \tag{4.10}$$

where c_{SOS} is a scaling factor (usually 1.3) and $E^{(2)}_{OS}$ is the opposite spin component of the MP2 energy:

$$E^{(2)}_{OS} = -\sum_{ia}\sum_{jb} \frac{(ia|jb)^2}{\epsilon_a + \epsilon_b - \epsilon_i - \epsilon_j} = -\sum_{ia}\sum_{jb} \frac{(ia|jb)^2}{\Delta^{ab}_{ij}}. \tag{4.11}$$

The energy denominator Δ^{ab}_{ij} in Equation (4.11) can be rewritten using the Laplace transformation $1/x = \int_0^\infty e^{-xt}dt$ so that [62, 63]:

$$E^{(2)}_{OS} = -\int_0^\infty dt \sum_{ia}\sum_{jb} (ia|jb)^2 e^{-t\Delta^{ab}_{ij}}. \tag{4.12}$$

The integral over t in Equation (4.12) can be approximated by a numerical quadrature. Considering N_q quadrature points $E^{(2)}_{OS}$ can be written as:

$$\begin{aligned} E^{(2)}_{OS} &= -\sum_{q}^{N_q}\sum_{ia}\sum_{jb} w_q (ia|jb)^2 e^{-t_q\Delta^{ab}_{ij}} \\ &= -\sum_{q}^{N_q}\sum_{ia}\sum_{jb} (\overline{ia|jb})^2 \end{aligned} \tag{4.13}$$

where $(\overline{ia|jb})$ represents an ERI calculated over scaled molecular orbitals defined as:

$$\overline{\psi_a}^q = w_q^{\frac{1}{8}} \psi_a e^{-\frac{1}{2}t_q\epsilon_a} \tag{4.14}$$

$$\overline{\psi_i}^q = w_q^{\frac{1}{8}} \psi_i e^{\frac{1}{2}t_q\epsilon_i}. \tag{4.15}$$

The Laplace SOS-MP2 energy can thus be evaluated employing a slightly modified version of a canonical MP2 algorithm at the price of $O(N^5)$ operations. However, the

introduction of the RI approximation, in this case, allows for a reduction of the formal computational effort to $O(N^4)$ [127]. In fact the integrals $(\overline{ia|jb})$ within RI can be approximated as:

$$(\overline{ia|jb})_{RI} = \sum_P B_P^{\overline{ia}} B_P^{\overline{jb}} \tag{4.16}$$

$$B_P^{\overline{ia}} = w_q^{\frac{1}{4}} B_P^{ia} e^{\frac{1}{2} t_q (\epsilon_i - \epsilon_a)} \tag{4.17}$$

substituting equation Equation (4.16) into Equation (4.13) and carrying out the two independent summations over ia and jb gives:

$$\begin{aligned} E_{OS}^{(2)} &= -\sum_q^{N_q} \sum_{ia} \sum_{jb} \sum_{PR} B_P^{\overline{ia}} B_P^{\overline{jb}} B_R^{\overline{ia}} B_R^{\overline{jb}} \\ &= -\sum_q^{N_q} \sum_{PR} \overline{Q}_{PR} \overline{Q}_{PR} \\ &= -\sum_q^{N_q} \mathrm{Tr}(\overline{\mathbf{Q}}\, \overline{\mathbf{Q}}^{\mathbf{T}}). \end{aligned} \tag{4.18}$$

The $\overline{\mathbf{Q}}$ matrix is a $N_a \times N_a$ matrix calculated as:

$$\overline{Q}_{PR} = \sum_{ia} B_P^{\overline{ia}} B_R^{\overline{ia}}. \tag{4.19}$$

This means that, for each quadrature point q, the calculation of the associated $\overline{\mathbf{Q}}$ requires only $O(ovN_a^2)$ operations. The overall computational effort is thus $O(N_q ovN_a^2)$. Since the number of quadrature points is independent of system size the RI-Laplace-SOS-MP2 method scales $O(N^4)$. The quadrature points can be found with a minimax approximation [149, 150] and as few as $6-8$ points can yield micro-Hartree accuracy.

4.2.4 RI Direct Random Phase Approximation Correlation Energy Method

The RPA correlation energy [4] is given as the difference between the zero point energy of two harmonic oscillator excitation problems for which the first includes correlated ground state (RPA) and the second not (configuration interaction singles CIS) [34–36]:

$$E_c^{\mathrm{RPA}} = \frac{1}{2} \sum_i (\omega_i - \nu_i) = \frac{1}{2} \mathrm{Tr}(\omega - \mathbf{A}) \tag{4.20}$$

where ω is the diagonal matrix of the positive RPA excitation energies that can be obtained from the non-Hermitian eigenvalue problem,

$$\begin{pmatrix} \mathbf{A} & \mathbf{B} \\ -\mathbf{B} & -\mathbf{A} \end{pmatrix} \begin{pmatrix} \mathbf{X} & \mathbf{Y} \\ \mathbf{Y} & \mathbf{X} \end{pmatrix} = \begin{pmatrix} \mathbf{X} & \mathbf{Y} \\ \mathbf{Y} & \mathbf{X} \end{pmatrix} \begin{pmatrix} \omega & \mathbf{0} \\ \mathbf{0} & -\omega \end{pmatrix}. \tag{4.21}$$

Within the direct-RPA (dRPA) approach, that is RPA without including exchange contributions [36, 43], the orbital rotation Hessian matrices **A** and **B** are defined as:

$$(A-B)_{ia,jb} = (\epsilon_a - \epsilon_i)\delta_{ij}\delta_{ab} \tag{4.22}$$
$$(A+B)_{ia,jb} = (\epsilon_a - \epsilon_i)\delta_{ij}\delta_{ab} + 2(ia|jb) \tag{4.23}$$

here all matrices have dimension $ov \times ov$. It is known from time dependent density functional theory that Equation (4.21) can be transformed to a Hermitian problem [151]:

$$\mathbf{MZ} = \mathbf{Z}\boldsymbol{\omega}^{\mathbf{2}}, \quad \mathbf{Z}^{\mathbf{T}}\mathbf{Z} = \mathbf{1} \tag{4.24}$$

where

$$\mathbf{M} = (\mathbf{A}-\mathbf{B})^{1/2}(\mathbf{A}+\mathbf{B})(\mathbf{A}-\mathbf{B})^{1/2}. \tag{4.25}$$

It can be finally shown that Equation (4.20) can be rewritten in term of the square root of **M** as [35]:

$$E_c^{\mathrm{dRPA}} = \frac{1}{2}\mathrm{Tr}(\mathbf{M}^{1/2} - \mathbf{A}). \tag{4.26}$$

According to Equation (4.26) a straightforward approach for computing E_c^{dRPA} implies the storage and calculation of $\mathbf{M}^{1/2}$ and so requires $O(N^4)$ memory and $O(N^6)$ computational effort.

However, the dRPA correlation energy within the RI approximation $E_c^{\text{RI-dRPA}}$ can be expressed in term of a frequency integral [43]

$$E_c^{\text{RI-dRPA}} = \frac{1}{2}\int_{-\infty}^{+\infty}\frac{d\omega}{2\pi}\mathrm{Tr}(\ln(\mathbf{1}+\mathbf{Q}(\omega)) - \mathbf{Q}(\omega)). \tag{4.27}$$

Here the frequency dependent matrix $\mathbf{Q}(\omega)$ has size $N_a \times N_a$ and is given by $\mathbf{Q}(\omega) = 2\mathbf{B}^T\mathbf{G}(\omega)\mathbf{B}$ where $\mathbf{G}(\omega)$ is a diagonal $ov\times ov$ matrix with elements $G_{ia,ia}(\omega) = (\epsilon_a - \epsilon_i)((\epsilon_a - \epsilon_i)^2 + \omega^2)^{-1}$. For a give ω, the computation of the integrand function in Equation (4.27) requires $O(N^4)$ operations. The integral of Equation (4.27) can be efficiently calculated by Clenshaw-Curtis numerical quadrature [152] and usually $30-40$ quadrature points are enough for micro-Hartree accuracy.

Thus the introduction of the resolution of the identity (RI) approximation to the MO-ERIs and the frequency integration techniques for computing $E_c^{\text{RI-dRPA}}$ lead to a reduction of the computational cost to $O(N^4 N_q)$ and $O(N^3)$ storage only, where N_q is the number of points used in the numerical quadrature of the integral in Equation (4.27).

4.2.5 RI Gaussian and Plane-Waves Method

The Gaussian and Plane-Waves (GPW) method has been shown to be an efficient approach for computing ERIs especially when periodic boundary conditions are considered [44]. The basis of the GPW approach for computation of the ERIs is the direct formulation of the half transformed integrals of the type $(ia|\lambda\sigma)$ in term of the electrostatic potential

v^{ia} of the pair density ρ^{ia}

$$\begin{aligned}(ia|\lambda\sigma) &= \int\int \psi_i(\vec{r}_1)\psi_a(\vec{r}_1)\frac{1}{\vec{r}_{12}}\phi_\lambda(\vec{r}_2)\phi_\sigma(\vec{r}_2)d\vec{r}_1 d\vec{r}_2 \\ &= \int \left[\int \frac{\rho^{ia}(\vec{r}_1)}{\vec{r}_{12}}d\vec{r}_1\right]\phi_\lambda(\vec{r}_2)\phi_\sigma(\vec{r}_2)d\vec{r}_2 \\ &= \int v^{ia}(\vec{r}_2)\phi_\lambda(\vec{r}_2)\phi_\sigma(\vec{r}_2)d\vec{r}_2. \end{aligned} \tag{4.28}$$

The form of the last equation is essentially identical to the one used in the GPW method[86] to compute matrix elements of the Hartree potential. Thus, the highly efficient implementation of that operation in CP2K[90] can be directly used and we refer to Ref. [91] for a detailed discussion.

Within the RI approximation, two types of ERIs have to be computed, the two center $(P|Q)$ and three center $(ia|P)$. Focusing on three center case, they are computed, Equation (4.7), starting from the integrals over AOs that are subsequently transformed with the two matrices $\mathbf{C}$ and $\mathbf{L^{-1}}$. Employing the GPW method, Equation (4.28), the index transformation over the auxiliary basis can be avoided, since it is possible to directly compute half transformed integrals for an associated density ρ as

$$\begin{aligned}(\mu\nu|P) &= \sum_R (\mu\nu|R)L^{-1}_{PR} \\ &= \int\int \phi_\mu(\vec{r}_1)\phi_\nu(\vec{r}_1)\frac{1}{\vec{r}_{12}}\left[\sum_R \phi_R(\vec{r}_2)L^{-1}_{PR}\right]d\vec{r}_1 d\vec{r}_2 \\ &= \int \phi_\mu(\vec{r}_1)\phi_\nu(\vec{r}_1)\left[\int \frac{\rho^P(\vec{r}_2)}{\vec{r}_{12}}d\vec{r}_2\right]d\vec{r}_1 \\ &= \int \phi_\mu(\vec{r}_1)\phi_\nu(\vec{r}_1)v^P(\vec{r}_1)d\vec{r}_1. \end{aligned} \tag{4.29}$$

The same approach holds for the $(P|Q)$ integrals with the only difference that the potential is calculate from the density associated to a single Gaussian auxiliary basis function.

Of central importance in GPW is then the representation of the density on a regular grid, which is equivalent to an expansion of the density in an auxiliary basis of plane waves (PW). The expansion is given by

$$\rho^P(\vec{R}) \approx \frac{1}{\Omega}\sum_{|\vec{G}|\leq G_c}\rho^P(\vec{G})e^{i\vec{G}\cdot\vec{r}} \tag{4.30}$$

where the sum over the reciprocal lattice vectors $\vec{G}$ is determined by the resolution of the grid. $\rho^P(\vec{G})$ are the Fourier coefficients of the density, and Ω is the volume of the simulation cell. Conventionally, the resolution of the grid is specified as the energy cutoff $\frac{1}{2}G_c^2$ that limits the kinetic energy of the PWs. Fast Fourier transforms (FFTs) efficiently change representation between real space ($\rho^P(\vec{R})$) and reciprocal space ($\rho^P(\vec{G})$). In particular, for a grid with S grid points, the transformation can be performed in linear scaling time ($O(S\log S)$). In reciprocal space, it becomes straightforward to solve the Poisson equation for the potential v^P

$$v^P(\vec{G}) = \frac{4\pi}{G^2}\rho^P(\vec{G}) \tag{4.31}$$

and an additional back FFT (FFT^{-1}) will yield the potential in real space. For $\vec{G} = 0$ the value of the potential is set to zero, this corresponds to a constant shift to the potential (or redefinition of the vacuum level) [153]. Thanks to the orthogonality of the occupied-virtual orbitals this shift has no influence on the final value of the $(ia|P)$ integrals [92].

Once the potential v^P is available on a regular real space grid, the numerical integration over the basis functions is performed by summing the product of the value of the potential and the primitive Gaussian functions (PGFs) over the grid points. For a given $|P)$, all matrix elements that are non-zero within a given threshold (ϵ_{grid}) can be obtained in linear scaling time. A further gain in efficiency is obtained by employing a multi-grid technique that represents the potential v^P on grids with increasingly coarser grid spacing. The accuracy of the multi-grid scheme is fixed by specifying a relative cutoff (E^{rel}_{cut}) that specifies the E_{cut} of the grid that will be employed for a primitive Gaussian function (PGF) with exponent 1.0.

Finally, $(\mu\nu|P)$ integrals are transformed into MO ERIs using (sparse) matrix multiplication. Introducing for a given $|P)$ vector the matrix of ERIs $\mathbf{S}^P$ ($S^P_{\mu\nu} = (\mu\nu|P)$), the ia elements of the matrix B^{ia}_P are obtained by two matrix multiplication as $\mathbf{C}_o^\dagger \mathbf{S}^P \mathbf{C}_v$, where $\mathbf{C}_o$ and $\mathbf{C}_v$ represent the coefficient matrices of the occupied and virtual orbitals. The multiplication by $\mathbf{C}_o$ can exploit the sparsity of $\mathbf{S}^P$, implying an $O(no)$ scaling per $|P)$ vector, while the final multiplication can not exploit sparsity and is asymptotically dominant, scaling as $O(onv)$. The thresholding in the sparse matrix multiplication is enforced using a threshold $\epsilon_{filter} \approx \epsilon_{grid}$.

As we have shown for the MP2-GPW method [44], the overall accuracy of the MP2 energy can be well controlled, and is on the order of $10^{-7} - 10^{-8}$ a.u. per heavy atom for $E_{cut} = 300$ Ry, $E^{rel}_{cut} = 50$ Ry, $\epsilon_{filter} = \epsilon_{grid} = 10^{-8}$.

4.3 Parallel Implementation of the RI-GPW methods

In this section the algorithms and the parallelization strategies for the methods introduced previously are presented and discussed in detail. The algorithms are split in two steps, the first deals with the computation of the ERIs $(ia|P)$, Section 4.3.1, and is in common for all methods, the second is specific to the type of correlation energy calculated (Section 4.3.2 Section 4.3.3 Section 4.3.4).

4.3.1 GPW Calculation of the $(ia|P)$ ERIs

The pseudocode for the parallel algorithm for computing the $(ia|P)$ ERIs with the GPW approach is presented in Figure 4.1, while its main features are summarized in Table 4.1, the rest of the section discusses this figure in detail.

The presented algorithm follows closely the integral computation part of our MP2-GPW method[44] with the main difference that all computation that were based on ia occupied-virtual pairs now are performed for auxiliary basis related quantities. The parallelization is achieved with a multi-level hybrid OpenMP/MPI scheme, and a careful process layout. The first level of parallelization corresponds to distributing the work performed for a single given auxiliary basis function ϕ_P or vector $|P) = \sum_R \phi_R L^{-1}_{PR}$. The

Assign each process its coordinate (n_P, n_w)

Create ranges $[P_{start}^{n_P}, P_{end}^{n_P}]$, $[a_{start}^{n_w}, a_{end}^{n_w}]$

Loop over P auxiliary basis functions $(P_{start}^{n_P} \leq P \leq P_{end}^{n_P})$

Calculate density $\rho^P(\vec{R}) = \phi_P(\vec{R})$ on the real space grid

Transfer $\rho^P(\vec{R}) \rightarrow \rho^P(\vec{G})$: $\rho^P(\vec{G}) = \mathbf{FFT}[\rho^P(\vec{R})]$

Solve Poisson's Equation: $\rho^P(\vec{G}) \rightarrow v^P(\vec{G})$

Transfer $v^P(\vec{G}) \rightarrow v^P(\vec{R})$: $v^P(\vec{R}) = \mathbf{FFT^{-1}}[v^P(\vec{G})]$

Integrate potential in real space: $I_Q^P = \int v^P(\vec{R})\phi_Q(\vec{R})d\vec{R}$ (all AUX Q)

Store I_Q^P

End P Loop

Redistribute I_Q^P integrals into $(Q|P)$ parallel distributed matrix

Cholesky decomposition of $(Q|P) = \mathbf{LL^T}$ (SCALAPACK)

Triangular inversion of $\mathbf{L} \rightarrow \mathbf{L^{-1}}$

Collect and store $\mathbf{L^{-1}}$ rows $M_R^P = L_{PR}^{-1}$ (all AUX R, $P_{start}^{n_P} \leq P \leq P_{end}^{n_P}$)

Loop over P auxiliary basis functions $(P_{start}^{n_P} \leq P \leq P_{end}^{n_P})$

Calculate density $\rho^P(\vec{R}) = \sum_R \phi_R(\vec{R})M_R^P$ on the real space grid

Transfer $\rho^P(\vec{R}) \rightarrow \rho^P(\vec{G})$: $\rho^P(\vec{G}) = \mathbf{FFT}[\rho^P(\vec{R})]$

Solve Poisson's Equation: $\rho^P(\vec{G}) \rightarrow v^P(\vec{G})$

Transfer $v^P(\vec{G}) \rightarrow v^P(\vec{R})$: $v^P(\vec{R}) = \mathbf{FFT^{-1}}[v^P(\vec{G})]$

Integrate potential in real space: $(\mu\nu|P) = S_{\mu\nu}^P = \int \phi_\mu(\vec{R})\phi_\nu(\vec{R})v^P(\vec{R})d\vec{R}$ (all μ,ν)

Index transformation $\mathbf{V} = (\mathbf{C}_o^\dagger \mathbf{S}^P)\mathbf{C}_v$

Redistribute and store $B_{ia}^P = V_{ia}$ (all i, $a_{start}^{n_w} \leq a \leq a_{end}^{n_w}$)

End P Loop

Figure 4.1: Pseudocode of the parallel algorithm for computing the $\mathbf{B}$ matrix of ERIs $(ia|P)$ with the GPW approach.

second level of parallelization corresponds to a distribution of these nearly independent calculations. The N_p processes available in total are therefor split in N_G groups, each group working on a given ϕ_P or $|P)$ and each consisting of N_w processes ($N_p = N_G N_w$). The first level of parallelization is complicated, involving parallel FFTs, halo-exchanges, and sparse matrix multiplications over N_w processes. However, this level is readily available, as it corresponds to the standard parallelization scheme for DFT calculations in CP2K [91]. The second level of parallelization is more straightforward, since it only requires a few steps of inter-group redistribution of two center ERIs $(Q|P)$ in order to calculate $\mathbf{L^{-1}}$.

The total work load for the integral computation is distributed by splitting the total number N_a of auxiliary basis function into N_G ranges $[P_{start}^{n_P}, P_{end}^{n_P}]$, each of them labeled with a given n_P coordinate, and assigned to the corresponding group. Additionally, each of the N_w processes within a group is given an index n_w, so that a processes is uniquely identified by its coordinate (n_P, n_w). Finally, the a virtual index is split in N_w ranges $[a_{start}^{n_w}, a_{end}^{n_w}]$, while a splitting of the occupied i is not necessary.

Table 4.1: Features of the parallel algorithm for computing the **B** matrix of ERIs $(ia|P)$ with the GPW approach expressed as "order of" the calculation parameters. n and N_a number of primary and auxiliary basis functions, o and v number of occupied and virtual orbitals, S grid size, N_G and N_w number of groups and group size, N_p number of processes. N_G, N_w and N_p are related by $N_p = N_G N_w$. The notation for the individual step is taken from the algorithm in Figure 4.1.

	Memory	Execution Time
Generation of $\mathbf{M^P}$ vectors:		
ρ^P calculation	$\frac{S}{N_w}$	$\frac{N_a S}{N_p}$
FFT and Poisson solver	$\frac{S}{N_w}$	$\frac{N_a S \log(S)}{N_p}$
I_Q^P calculation (v^P integration)	$\frac{N_a^2}{N_G}$	$\frac{N_a^2}{N_p}$
I_Q^P redistribution into $(P\|Q)$	$\frac{N_a^2}{N_G}$	N_a^2
Cholesky decomposition of $(P\|Q)$	$\frac{N_a^2}{N_p}$	$\frac{N_a^3}{N_p}$
Generation of $\mathbf{L^{-1}}$	$\frac{N_a^2}{N_p}$	$\frac{N_a^3}{N_p}$
Collect $\mathbf{M^P}$ vectors	$\frac{N_a^2}{N_G}$	N_a^2
Generation of $(ia\|P)$ integrals (**B** matrix):		
ρ^P calculation	$\frac{S}{N_w}$	$\frac{N_a S}{N_p}$
FFT and Poisson solver	$\frac{S}{N_w}$	$\frac{N_a S \log(S)}{N_p}$
$S_{\mu\nu}^P$ calculation (v^P integration)	$\frac{S+n}{N_w}$	$\frac{N_a n}{N_p}$
1^{st} index transformation	$\frac{on}{N_w}$	$\frac{N_a on}{N_p}$
2^{nd} index transformation	$\frac{ov}{N_w}$	$\frac{N_a ovn}{N_p}$
Redistribution and Storage **B**	$\frac{ovN_a}{N_p}$	$\frac{N_a ov}{N_G}$

The algorithm is split in two parts, the first related to the calculation of $\mathbf{L^{-1}}$ the second to the computation of the **B** matrix of $(ia|P)$. In the two steps there are common features, in fact, in both cases, the basic structure of the GPW machinery can be recognized, consisting of a density $\rho(\vec{R})$ calculation on the real space grid, the computation of the associated potential $v(\vec{R})$ and the final integration of $v(\vec{R})$ over Gaussian elements of the basis. The difference in the two cases is that for the computation of $\mathbf{L^{-1}}$, $\rho(\vec{R})$ is related to a single auxiliary basis function and the associated potential is integrand over the auxiliary basis set functions, while for the computation of the $(ia|P)$ ERIs $\rho(\vec{R}) = \sum_R \phi_R L_{PR}^{-1}$ and the integration is performed over pair elements of the primary basis $(\mu\nu)$.

At the end of the first loop over auxiliary basis elements, each group stores a slice of the $(Q|P)$ matrix (labeled as I_Q^P in the pseudocode in Figure 4.1) comprised all Q and $P \in [P_{start}^{n_P}, P_{end}^{n_P}]$. In order to perform efficiently the Cholesky decomposition $(Q|P) = \mathbf{L}\mathbf{L}^{\mathbf{T}}$ and subsequently the $\mathbf{L}$ inversion, the I_Q^P are redistributed into a 2D layout of a parallel distributed matrix as used in SCALAPACK. Once the $\mathbf{L^{-1}}$ matrix is computed a similar redistribution procedure is again performed in order to collect the rows of $\mathbf{L^{-1}}$ ($M_R^P = L_{PR}^{-1}$, all R, $P \in [P_{start}^{n_P}, P_{end}^{n_P}]$) necessary for the next step. The time spent for these two redistribution steps is not scaling well with the number of processes, nevertheless the communication effort grows only $O(N^2)$ and the associated time, for the benchmark calculations performed so far, is negligibly small.

The second loop over auxiliary basis vector $|P)$ allows for the final computation of the $(ia|P)$ integrals. The time for the calculation of the $\mathbf{S}^P$ matrix is linear scaling with the system size for a given P, since only pairs of overlapping Gaussians $(\mu\nu|$ need to be considered, and only a finite number of grid points within a spherical region around the center of the PGF is required. This implies that the total time for the integration of the potential in the RI-GPW method is $O(N^2)$, which is a great reduction compared to the canonical GPW-MP2 algorithm, where the corresponding task is $O(N^3)$.

Moreover, another advantage of the RI-GPW method, compared to standard RI integral implementation, is that only two integral transformations from AO to MO basis are required. The third transformation, which in conventional RI-MP2 implementation is performed as $(ia|P) = \sum_R (ia|R) L_{PR}^{-1}$, and requires $O(ovN_a^2)$ operations, is no longer needed.

At the end of the RI-GPW integral algorithm each process stores the matrix elements B_{ia}^P for all i, $P \in [P_{start}^{n_P}, P_{end}^{n_P}]$ and $a \in [a_{start}^{n_w}, a_{end}^{n_w}]$.

4.3.2 RI-MP2 Method

Once the integrals $(ia|P)$ are available the calculation of the RI-MP2 energy, in a serial algorithm, is straightforward, since the only tasks are related to the $(ia|jb)$ integral generation, Equation (4.5), and the energy accumulation, Equation (4.8).

In a parallel algorithm the main complication is introduced by the distributed storage of the $(ia|P)$ integrals. In particular, due to the features of the of the RI-GPW algorithm, previously described, each process stores elements B_{ia}^P for all occupied i, $P \in [P_{start}^{n_P}, P_{end}^{n_P}]$ and $a \in [a_{start}^{n_w}, a_{end}^{n_w}]$. Since N_w is usually small compared to the total number of processes, the virtual index a is distributed over a small number of MPI tasks within the group G while the auxiliary index P is distributed over the large amount of N_G groups.

In our RI-MP2 parallel algorithm, the $(ia|jb)$ integral generation proceeds as follow: first the independent ij pairs $(i \leq j)$ are statically distributed over the N_G groups, for each ij pair, the full range of the auxiliary index P is collected on a local buffer from all other groups, while keeping the virtual index distribution within the group, finally the $(ia|jb)$ integrals are generated for the actual ij pair in a matrix-multiplication fashion (Equation (4.5)) requiring only a small amount of communication within the group. Once the $(ia|jb)$ are available, they are accumulated into the MP2 energy according to Equation (4.8), requiring an additional negligible amount of communication within the group.

With this choice, the main source of inter-group communication in the parallel algorithm

Check available memory [Mem]

According to [Mem] define the replication group R of size N_r

According to [Mem] define the batch size B_S

Split the N_G groups into N_R subgroups ($N_R = N_G/N_r$)

Assign to each group G its coordinate in the subgroup n_R

Collect B_{ia}^P from all other processes with same coordinate (n_R, n_w)

Distribute IJ batches over the N_G groups ($I \leq J$, batch size of I and J given by B_S)

Loop over IJ batches ($IJ \in$ my_G^{IJ})

Collect $A_{aP}^i = B_{ia}^P$ and $E_{aP}^j = B_{ja}^P$, ($i \in I$, $j \in J$, all P, $a \in [a_{start}^{n_w}, a_{end}^{n_w}]$)

from all other processes in R with my same n_w coordinate

Loop over ij ($i \in I$, $j \in J$)

$(ia|jb) = I_{ab} = \sum_P A_{aP}^i E_{bP}^j$ $(a, b \in [a_{start}^{n_w}, a_{end}^{n_w}])$

Loop over n'_w, $n'_w \in G$

Collect E_{bP}^j, $(b \in [a_{start}^{n'_w}, a_{end}^{n'_w}])$

$(ia|jb) = I_{ab} = \sum_P A_{aP}^i E_{bP}^j$, $(a \in [a_{start}^{n_w}, a_{end}^{n_w}], b \in [a_{start}^{n'_w}, a_{end}^{n'_w}])$

End n'_w Loop

$E^{(2)} = E^{(2)} - (2 - \delta_{ij}) \sum_{ab} I_{ab}(2I_{ab} - I_{ba})/(\Delta_{ij}^{ab})$, $(a, b \in [a_{start}^{n_w}, a_{end}^{n_w}])$

Loop over n'_w, $n'_w \in G$

Collect I'_{ba}, $(b \in [a_{start}^{n_w}, a_{end}^{n_w}], a \in [a_{start}^{n'_w}, a_{end}^{n'_w}])$

$E^{(2)} = E^{(2)} - (2 - \delta_{ij}) \sum_{ab} I_{ab}(2I_{ab} - I'_{ba})/(\Delta_{ij}^{ab})$, $(a \in [a_{start}^{n_w}, a_{end}^{n_w}], b \in [a_{start}^{n'_w}, a_{end}^{n'_w}])$

End n'_w Loop

End ij Loop

End IJ Loop

Global summation of $E^{(2)}$

Figure 4.2: Pseudocode of the parallel algorithm for computing the RI-MP2 energy from the ERIs $(ia|P)$.

is related to the redistribution of the B_{ia}^P integrals, required for each ij pair. In order to perform this task efficiently, so that the required communication is reduced increasing the number of processes, a scheme involving replication of the B_{ia}^P into subgroup and batch communication for the ij pairs, has been designed. Both replication and batching allow to improve the efficiency in communication since, in the first case, the number of processes that have to communicate with each other is reduced, in the latter because the number of messages is reduced. The available memory is thus used to reduce communication.

The pseudocode of the RI-MP2 energy calculation algorithm is sketched in Figure 4.2 and the main features of the algorithm are summarized in Table 4.2. In a first stage, according to the available memory the replication group size N_r is defined and integrals are replicated among processes. As shown in Table 4.2 the time involved in this step increases linearly with the logarithm of the number of replication group N_R, while it decreases when the number of processes ($N_G N_w$) is increased. We have observed that a

Table 4.2: Features of the parallel algorithm for computing the correlation energy with the different methods expressed as "order of" the calculation parameters. The meaning of the different symbols is referred to Table 4.1 with the only addition of: N_R and N_r that represent the number of replication/integration group and the size of the replication/integration group ($N_G = N_R N_r$), B_S that is the batch size for ij in RI-MP2, and N_q that is the number of quadrature points used for the numerical quadrature in Laplace-RI-SOS-MP2 and RI-dRPA. The notation of the individual step is referred to the different algorithms.

	Memory	Execution Time
RI-MP2 algorithm (Section 4.3.2):		
Replication of $(ia\|K)$ into R groups	$\frac{ovN_a}{N_w N_r}$	$\frac{\log_2(N_R)}{N_G}\frac{ovN_a}{N_w}$
$(ia\|jb)$ integral generation	$\frac{2B_S vN_a}{N_w}$	$\frac{o^2v^2N_a}{N_p}$
MP2 energy accumulation	$\frac{v^2}{N_w}$	$\frac{o^2v^2}{N_p}$
Communication	$\frac{2B_S vN_a}{N_w}$	$\frac{o^2vN_a}{N_p B_S}$
RI-dRPA algorithm (Section 4.3.3):		
Creation of $\mathbf{B}$ matrix	$\frac{ovN_a}{N_w N_r}$	$\frac{N_p\log_2(N_R)+(N_wN_r-1)}{N_p}\frac{ovN_a}{N_rN_w}$
Calculation of $\mathbf{B'}(\omega) = \mathbf{G}(\omega)\mathbf{B}$	$\frac{ovN_a}{N_w N_r}$	$\frac{N_q ovN_a}{N_p}$
Calculation of $\mathbf{Q}(\omega) = 2\mathbf{B^T B'}(\omega)$	$\frac{ovN_a}{N_w N_r}$	$\frac{N_q ovN_a^2}{N_p}$
Calculation of $\mathrm{Tr}[\ln(\mathbf{Q}(\omega)+\mathbf{1}) - \mathbf{Q}(\omega)]$	$\frac{N_a^2}{N_w N_r}$	$\frac{N_q N_a^3}{N_p}$
Laplace-RI-SOS-MP2 algorithm (Section 4.3.4):		
Calculation of $\mathbf{B'}$	$\frac{ovN_a}{N_w N_r}$	$\frac{N_q ovN_a}{N_p}$
Calculation of $\overline{\mathbf{Q}} = \mathbf{B^T B'}$	$\frac{ovN_a}{N_w N_r}$	$\frac{N_q ovN_a^2}{N_p}$
Calculation of $\mathrm{Tr}[\overline{\mathbf{Q}}\,\overline{\mathbf{Q}}^{\mathbf{T}}]$	$\frac{N_a^2}{N_w N_r}$	$\frac{N_q N_a^2}{N_p}$

ratio N_R/N_G of ~ 0.1 is usually a good compromise between the time necessary for the replication and the gain in communication in the subsequent phase.

Once the R subgroups have been created and the B_{ia}^P integrals replicated, the maximum possible batch size B_S is defined based on the available memory per process. The total number of IJ batches ($I \leq J$) are then distributed statically over the groups. For load balancing reasons, the number of IJ batches is restricted to be a multiple of the number of groups N_G and the remaining ij single pairs are again statically distributed over groups.

At this point, each group loops over its assigned IJ batches and, collects the $A_{aP}^i = B_{ia}^P$ and $E_{aP}^j = B_{ja}^P$ integrals from the other members of the replication group R. The index ranges are $i \in I$, $j \in J$, all P, $a \in [a_{start}^{n_w}, a_{end}^{n_w}]$. This means that, if $N_w = 1$ the

index a spans the full range of virtual orbitals and all the subsequent operations are performed locally. If $N_w > 1$, the E^j_{aP} integrals have to be exchanged in a parallel matrix-multiplication-like fashion for the generation of the $(ia|jb)$ integrals that are then contracted into the MP2 energy, requiring an additional amount of in-group communication. As stated before, the in-group communication, which involves few processes, is usually negligible compared to the total communication time, and it is thus not taken into account in Table 4.2.

The total number of times the communication routine is invoked is $O\left(\frac{o^2}{B_S^2 N_G}\right)$ while the time required for each event of communication (considered to be proportional to the message sizes) is $O\left(\frac{vN_aB_S}{N_w}\right)$. This makes the total time spent in communication $O\left(\frac{o^2vN_a}{N_pB_S}\right)$, that is, communication is reduced when increasing the number of processes $N_p = N_wN_G$ and increasing the batch size B_S.

In the presented algorithm, the time determining step is the $(ia|jb)$ integral generation that is essentially a local matrix-matrix multiplication. This allows to fully exploit the performance of highly optimized routines, such as DGEMM. Moreover, this step can be further accelerated by employing a hybrid implementation that utilizes graphics processing units (GPUs).

4.3.3 RI-dRPA Method

The presented dRPA correlation energy implementation is based on the method developed by Eshuis *et al.* [43]. In this approach, the calculation of the integral in Equation (4.27) is accomplished by Clenshaw-Curtis numerical quadrature [152]. Here, only the main features of such a method are reported referring to the original paper for more details.

Given N_q quadrature points, the working expressions for the numerical quadrature of Equation (4.27) are:

$$F^C(\omega) = \frac{1}{2}\mathrm{Tr}[\ln(\mathbf{Q}(\omega) + \mathbf{1}) - \mathbf{Q}(\omega)] \tag{4.32}$$

$$E_c^{\text{RI-dRPA}} \simeq \sum_{q=1}^{N_q} \frac{a \cdot w_q}{2\pi} F^C(a \cdot \cot(t_q)) \tag{4.33}$$

$$t_q = \frac{q}{N_q}\frac{\pi}{2}, \quad q = 1, \ldots, N_q \tag{4.34}$$

$$w_q = \begin{cases} \frac{\pi}{N_q \sin^2 t_q}, & q = 1, \ldots, N_q - 1 \\ \frac{\pi}{2N_q \sin^2 t_q}, & q = N_q. \end{cases} \tag{4.35}$$

where a is a scaling parameter that ensures that the grid points are adaptively distributed over the spectrum of eigenvalues of matrix $\mathbf{M}$ (Equation (4.24) and Equation (4.25)). Following Ref. [43], the calculation of a can be performed with $O(N^2)$ computational cost, in our algorithm the bisection method is used instead of Newton-Raphson.

The pseudocode of the parallel algorithm for the calculation of $E_c^{\text{RI-dRPA}}$ is reported in Figure 4.3. The parallel algorithm is based on a two level work load distribution. The first level corresponds to the distribution of the work necessary for a given quadrature point q,

Check available memory [Mem]

According to [Mem] define the minimum size $N_r^{\min}$ of the integration group R

Define the size N_r and number $N_R = N_G/N_r$ of integration group, such that:

$N_r \geq N_r^{\min}$ and N_R is a divisor of the number of quadrature points N_q

Split the N_G groups into N_R subgroups

Assign to each integration group R its subset of quadrature points $\{q^{\text{my}}\}$

Create the matrix $\mathbf{B}$ of size $ov \times N_a$ distributed within the integration group R

Collect B_{ia}^P integrals from all other processes and fill $\mathbf{B}$ matrix

According to N_q calculate Clenshaw-Curtis weight $\{w_q\}$ and abscissa $\{t_q\}$

Calculate the scaling parameter a

Loop over q quadrature points ($q \in \{q^{\text{my}}\}$)

Calculate $\omega = a \cdot \cot(t_q)$

Calculate $\mathbf{B}'(\omega) = \mathbf{G}(\omega)\mathbf{B}$

Calculate $\mathbf{Q}(\omega) = 2\mathbf{B}^{\mathbf{T}}\mathbf{B}'(\omega)$

Calculate $F^C(\omega) = \frac{1}{2}\text{Tr}[\ln(\mathbf{Q}(\omega) + \mathbf{1}) - \mathbf{Q}(\omega)]$

$E_c^{\text{RI-dRPA}} = E_c^{\text{RI-dRPA}} + F^C(\omega) \cdot a \cdot w_q/(2\pi)$

End q Loop

Global summation of $E_c^{\text{RI-dRPA}}$ across the integration groups

Figure 4.3: Pseudocode of the parallel algorithm for computing the RI-dRPA energy from the ERIs $(ia|P)$ and Clenshaw-Curtis numerical quadrature of Equation (4.27).

the second to the distribution of the independent quadrature points over subgroups of processes (the integration groups R).

The second level of parallelization is straightforward, first, according to the available memory, the minimum size $N_r^{\min}$ for the integration group is defined. Then, the actual size N_r is obtained such that $N_r \geq N_r^{\min}$ and the number of integration group $N_R = N_G/N_r$ is an exact divisor of the total number of quadrature points N_q. In this way, each integration group R has an identically sized N_q/N_R subset of quadrature points $\{q^{\text{my}}\}$.

Once N_r is defined, the $\mathbf{B}$ matrix, defined in SCALAPACK format, has to be made available within the integration group R. This is accomplished in two steps. In the first step, the parallel distributed matrix $\mathbf{B}$, of size $ov \times N_a$, is allocated over the members of R, each process in R is identified with its coordinate n_r. Subsequently, the locally held data (the $(ia|P) = B_{ia}^P$ distributed as described in Section 4.3.1) are redistributed within R according to new ranges of $\mathbf{B}$. This step consists of in-group communication and requires $O\left(\frac{(N_w N_r - 1)}{N_w N_r}\frac{ov N_a}{N_p}\right)$ effort. The second step consists of inter-group communication, requiring replication of data across the integration groups. Since the integration groups have all the same size $(N_w N_r)$ and the matrix $\mathbf{B}$ is created in the same manner, it means that $\mathbf{B}$ retains the same structure for all groups, that is, all processes with the same coordinate n_r (belonging to different groups), have the same ranges for the rows and columns. Consequently, in the second step, only processes that have the same coordinate n_r communicate, replicating their local data across the integrations groups. This requires a

communication effort of $O\left(\frac{\log_2(N_R)}{N_w N_r} ovN_a\right)$, giving the total cost reported in table Table 4.2.

After the $\mathbf{B}$ matrix is created, the algorithm proceeds independently for each integration group. As a first task the matrix $\mathbf{B}'$ is calculated as $\mathbf{G}(\omega)\mathbf{B}$. Since $\mathbf{G}(\omega)$ is a $ov \times ov$ diagonal matrix with elements $G_{ia,ia}(\omega) = (\epsilon_a - \epsilon_i)((\epsilon_a - \epsilon_i)^2 + \omega^2)^{-1}$ the calculation of $\mathbf{B}'$ requires only $O\left(\frac{ovN_a}{N_w N_r}\right)$ operations for each quadrature point without inter-process communication.

The time determining step of the algorithm is the calculation of the $N_a \times N_a$ matrix $\mathbf{Q}(\omega)$ computed as $2\mathbf{B}^{\mathbf{T}}\mathbf{B}'(\omega)$. This task is performed as a standard parallel matrix-matrix multiplication using SCALAPACK. Consequently the computational effort per quadrature point is $O\left(\frac{ovN_a^2}{N_w N_r}\right)$ while the communication is expected to scale as $O(1/\sqrt{N_w N_r})$.

The last computationally demanding task is the calculation of $\mathrm{Tr}[\ln(\mathbf{Q}(\omega) + \mathbf{1})]$. This step can be efficiently carried out by considering the identity $\mathrm{Tr}[\ln \mathbf{A}] = \ln(\mathrm{Det}[\mathbf{A}])$, that is:

$$\mathrm{Tr}[\ln(\mathbf{Q}(\omega) + \mathbf{1})] = 2\sum_{i=1}^{N_a} \ln(U_{ii}) \tag{4.36}$$

where the $\mathbf{U}$ matrix is the Cholesky decomposition of $\mathbf{Q}(\omega) + \mathbf{1}$.

4.3.4 Laplace-RI-SOS-MP2 method

The Laplace-RI-SOS-MP2 algorithm is closely related to the RI-dRPA algorithm, in fact in both cases a numerical quadrature is required together with an $O(N^4)$ matrix-matrix multiplication step.

Contrary to the RI-dRPA case, optimal integration grids for the numerical quadrature in Laplace-transform MP2 method can be obtained relatively easily. In fact, the energy denominator Δ_{ij}^{ab} depends only on the occupied and virtual orbital energies such that $\Delta_{ij}^{ab} \in [E_{\min}, E_{\max}]$, where $E_{\min} = 2(\epsilon_{\mathrm{LUMO}} - \epsilon_{\mathrm{HOMO}})$ and $E_{\max} = 2(\epsilon_{\max} - \epsilon_{\min})$ being $\epsilon_{\max}$ and $\epsilon_{\min}$ the maximum and minimum value of the orbital energies respectively. This implies that, in order to reach reasonable accuracy ($10^{-5} - 10^{-6}$ Hartree) in the computation of the $E_{OS}^{(2)}$, in general, much less integration points are required compared to the RI-dRPA case. Different quadrature techniques such as Gauss-Laguerre, Gauss-Legendre [154], and least-squares (LS) quadrature [63] have been employed. In our implementation, the minimax approximation [149, 150] has been adopted. The minimax approach has some appealing features such as uniformity of the error along the whole range and error of alternating sign, leading to possible error cancellation. It has been shown that the minimax method in the Laplace-transform MP2 framework remarkably outperforms classical quadrature techniques, while giving comparable accuracy compared to the LS approach [155]. A detailed description of the implementation of the minimax quadrature scheme in CP2K is reported in Appendix A.

As shown in the pseudocode given in Figure 4.4, the algorithm for computing $E_{OS}^{(2)}$ proceeds in the same way as in the RI-dRPA case till the generation of the $\mathbf{B}$ matrix. When the $\mathbf{B}$ matrix has been replicated and redistributed within each integration group, the minimax parameters $\{w_q^*\}$ and $\{t_q^*\}$ are determined for the range $\Delta_R = E_{\max}/E_{\min}$, that is equivalent to find the minimax approximation of the $1/x$ function for $x \in [1, \Delta_R]$.

Check available memory [Mem]

According to [Mem] define the minimum size $N_r^{\min}$ of the integration group R

Define the size N_r and number $N_R = N_G/N_r$ of integration group, such that:

$N_r \geq N_r^{\min}$ and N_R is a divisor of the number of quadrature points N_q

Split the N_G groups into N_R subgroups

Assign to each integration group R its subset of quadrature points $\{q^{\text{my}}\}$

Create the matrix $\mathbf{B}$ of size $ov \times N_a$ distributed within the integration group R

Collect B_{ia}^P integrals from all other processes and fill $\mathbf{B}$ matrix

Calculate $E_{\min} = 2(\epsilon_{\text{LUMO}} - \epsilon_{\text{HOMO}})$ and $E_{\max} = 2(\epsilon_{\max} - \epsilon_{\min})$

Calculate $\Delta_R = E_{\max}/E_{\min}$

According to N_q and Δ_R calculate minimax weight $\{w_q^*\}$ and abscissa $\{t_q^*\}$

Scale minimax parameters $w_q = w_q^*/E_{\min}$, $t_q = t_q^*/E_{\min}$

Loop over q quadrature points ($q \in \{q^{\text{my}}\}$)

  Loop over local columns P

    Loop over local rows ia

      $B_{ia}^{P\prime} = B_{ia}^P \cdot \sqrt{w_q} \cdot \exp(t_q(\epsilon_i - \epsilon_a))$

    End ia Loop

  End P Loop

  Calculate $\overline{\mathbf{Q}} = \mathbf{B^T B'}$

  $E_{OS}^{(2)} = E_{OS}^{(2)} - \text{Tr}[\overline{\mathbf{Q}}\,\overline{\mathbf{Q}}^{\mathbf{T}}]$

End q Loop

Global summation of $E_{OS}^{(2)}$ across the integration groups

Figure 4.4: Pseudocode of the parallel algorithm for computing the Laplace-RI-SOS-MP2 energy from the ERIs $(ia|P)$ and the minimax approximation for the numerical quadrature in Equation (4.18).

These parameters are then scaled by $E_{\min}$, that is, the range of the approximation is shifted to $[E_{\min}, E_{\max}]$.

Each integration group R will perform all the tasks for its preassigned quadrature points $\{q^{\text{my}}\}$ in parallel within the members of R. The required tasks for a given quadrature point consist in updating the $\mathbf{B'}$ matrix with the actual weight w_q and abscissa t_q, perform the matrix-matrix multiplication $\overline{\mathbf{Q}} = \mathbf{B^T B'}$ and increment the $E_{OS}^{(2)}$ with the q^{th} contribution given by $\text{Tr}[\overline{\mathbf{Q}}\,\overline{\mathbf{Q}}^{\mathbf{T}}]$. The computational effort required for each of these individual step is reported in Table 4.2. While the update of $\mathbf{B'}$ and the $O(N^4)$ matrix multiplication are analogous to the corresponding steps in the RI-dRPA algorithm, the calculation of $\text{Tr}[\overline{\mathbf{Q}}\,\overline{\mathbf{Q}}^{\mathbf{T}}]$ is only $O(N^2)$ and is performed without inter-processes communication.

4.3.5 Hybrid CPU/GPU Implementations

For the methods presented in this chapter, the time determining step, from a computational complexity point of view, is a matrix-matrix multiplication. In the case of the parallel RI-MP2 algorithm it is a process-local matrix multiplication (DGEMM), while in the other two cases it is a parallel matrix multiplication (PDGEMM).

New accelerator hardware, such as graphics processing units (GPUs), can perform these operations efficiently, with a performance exceeding that of several traditional CPU codes both in time to solution as well as energy efficiency.

In the current GPU implementation, only the matrix multiplication step is performed on the GPU. This implies that the impact is limited to sufficiently large systems, where this part is dominating.

4.4 Benchmark Calculations

4.4.1 Computational Details

The EXX/RPA Formalism

The exact exchange (EXX) and random phase approximation correlation energies formalism (EXX/RPA) has been extensively applied to a large variety of systems including isolated molecules [7–10, 12, 42, 135, 138, 156, 157], solids [158–162], surfaces [160, 163–166] and van der Waals bonded crystals [167–169].

Within the framework of EXX/RPA formalism the total energy is given as:

$$\begin{aligned} E_{\mathrm{tot}}^{\mathrm{EXX/RPA}} &= E_{\mathrm{tot}}^{\mathrm{HF}} + E_{\mathrm{c}}^{\mathrm{RPA}} \\ &= \left(E_{\mathrm{tot}}^{\mathrm{DFT}} - E_{\mathrm{xc}}^{\mathrm{DFT}}\right) + E_{\mathrm{x}}^{\mathrm{EXX}} + E_{\mathrm{c}}^{\mathrm{RPA}} \end{aligned} \tag{4.37}$$

where the right-hand side terms of last equation are the DFT total energy, the exchange-correlation DFT energy, the exact exchange and the RPA correlation energy respectively. The sum of the first three terms is referred as the Hartree-Fock (HF) energy calculated employing the DFT orbitals, and in the following will be denoted simply as HF@DFT. The last term correspond to the RPA correlation energy computed using the DFT orbitals and orbital energies and will be referred as RPA@DFT.

The calculation of the $E_{\mathrm{tot}}^{\mathrm{EXX/RPA}}$ for a given system is thus performed by first converging the self consistent field (SCF) procedure with a given DFT method. Then the ground state single-particle wave functions and orbital energies are used as input to compute the EXX energy and the RPA correlation energy.

All EXX/RPA calculations reported here have been performed employing Kohn-Sham Perdew-Burke-Ernzerhof (PBE) [170] orbitals as input unless otherwise stated.

Basis Sets, Thresholds and Pseudopotentials

The RI-GPW methods as implemented in CP2K[90] have been employed for all calculations in this manuscript. The correlation energy calculations are based on pseudopotentials of the form suggested by Goedecker, Teter and Hutter (GTH) in Ref. [101] but specifically parameterized for Hartree-Fock (HF) and DFT calculations. In this way, core states do

not need to be represented and valence orbitals are smooth, as required by the GPW method. The same basis used in the previous GPW-MP2 chapter has been adopted [44]. They consist in valence-only correlation consistent type [102, 103] basis sets, generated for being used with these pseudopotentials. The basis sets have been labeled as cc-DZVP, cc-TZVP and cc-QZVP, denoting double, triple and quadruple zeta quality respectively. For each primary basis set, the relative auxiliary RI basis has been generated according to the procedure proposed by Weigend *et al.* [69] (see Appendix B for details). Using uncontracted Gaussian functions only, the size of the auxiliary basis is typically between 2.5 and 3 times larger than the corresponding primary basis and includes functions with angular momentum quantum number up to $(l_{max}+1)$, where l_{max} is the maximum $l-$quantum number of the primary basis. For example, in the case of oxygen, the auxiliary basis associated to the cc-TZVP basis is built with the following pattern of primitives (6s 5p 5d 3f 1g). The condition number of the $(P|Q)$ matrix, has been computed employing both normalized and unnormalized auxiliary basis functions. In the case of molecular crystals described with the cc-TZVP basis, the largest observed condition number is of the order of 10^8 and 10^{10} for the normalized and unnormalized case respectively. In the case of the LiH crystal, considering the $3\times3\times3$ cell, employing the cc-QZVP basis, the condition number is of the order of 10^{10} for both the normalize and unnormalized case. These different normalizations have shown not to give discrepancies for the computed cohesive energies.

The Hartee-Fock exchange calculations have been performed employing our robust Γ-point implementation [104, 105] that allows for stable calculations in the condensed phase [105, 106]. The Schwarz screening threshold for the HF calculations is in the range $10^{-8}-10^{-10}$, periodic calculations used a truncated Coulomb operator[105], using approximately half the length of the smallest edge (L) of the simulation cell as truncation radius (R_c), typically 5-6Å. We have verified that the effect on the MP2 energy due to the truncation is usually one order of magnitude smaller than the effect on the Hartree-Fock energy. The latter has been shown to converge exponentially[105], and for the benzene crystal, the error in the cohesive energy is 1kJ/mol and 0.2kJ/mol for an R_c of 5Å and 7Å, respectively. Note that, as we discussed in Ref. [44], the MP2 energy computed at the Γ-point displays a slower convergent (L^{-3}) size-effect.

The threshold for the SCF convergence was 10^{-6} or tighter. The PW cutoff for the HF or DFT part of the calculations was $E_{cut}=1200$ Ry to guarantee convergence of the exchange-correlation term, at small cost compared to the post SCF correlation energy calculations. The HF@DFT calculations have been performed retaining the pseudopotential of the DFT method. The correlation energy calculations employed high quality PW cutoffs of $E_{cut}=300$ Ry, $E_{cut}^{rel}=50$ Ry, $\epsilon_{filter}=10^{-7}$, and $\epsilon_{grid}=10^{-6}$, unless mentioned otherwise. Gas phase systems have been computed using cluster boundary conditions for solving the Poisson equation [95].

In order to validate the reported settings, calculations for the S22 set [130] have been performed for the RI-MP2 and RI-dRPA methods. The tested basis sets are cc-TZVP, cc-QZVP and a slightly modified cc-TZVP basis containing augmented functions (labeled as aug-cc-TZ). The obtained results for the S22 set show good agreement with the previously reported calculations for basis of comparable quality [130, 133, 171], for example, the mean absolute deviation, in the case of the aug-cc-TZ with respect to Dunning's augmented correlation consistent basis aug-cc-pVTZ is 0.1 kcal/mol for both RI-MP2 and RI-dRPA.

The HF and PBE pseudopotentials, all basis sets employed here, as well as the results of the calculations on the S22 set can be found in the supporting information of Ref. [131].

Geometries and Cohesive Energies

For all crystals, supercells have been generated by replicating the unit cell, so that the smallest edge was larger than 9Å, in order for the Γ-point approximation to be reasonable. The experimental geometries of the molecular crystals have been retrieved from the Cambridge Structural Database (CSD)[108]. The positions of the hydrogen atoms of these geometries have been further relaxed at the DFT/BLYP [109, 110] level employing the cc-TZVP basis set, unless otherwise stated. The main features of the structure of each crystal, together with the supercell used in the calculation and the CSD refcode, are reported in the supporting information of Ref. [131] (see also Figure 3.4 and Ref. [44]). The LiH crystal geometry is based on the experimental value of the lattice parameter ($a = 4.084$ Å) .

The counterpoise (CP) corrected cohesive energy per molecule at a given volume V has been computed as [107, 111]

$$E_{coh}^{CP}(V) = \frac{E_{supercell}(V)}{N_{mol}} - E_{mol}^{gas} - E_{mol+ghost}^{crystal}(V) + E_{mol}^{crystal}(V). \tag{4.38}$$

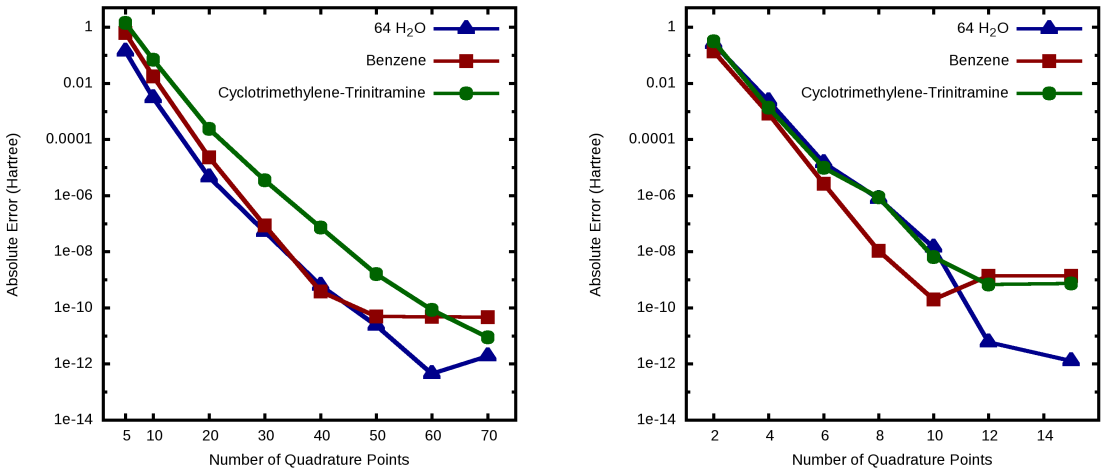
Here, N_{mol} is the number of molecules per supercell, $E_{supercell}(V)$ the total energy of the supercell, and $E_{mol+ghost}^{crystal}(V)$, $E_{mol}^{crystal}(V)$, and E_{mol}^{gas} the total energy of an isolated molecule in either the crystal geometry ($E_{mol+ghost}^{crystal}(V)$ and $E_{mol}^{crystal}(V)$) or a gas phase geometry (E_{mol}^{gas}). $E_{mol+ghost}^{crystal}(V)$, includes ghost atoms from the molecules of first coordination shell, while the gas phase geometries has been obtained by relaxation at the B3LYP/cc-TZVP level[110, 112, 113]. To assess the accuracy of computed cohesive energies, these values have been compared to the experimental sublimation enthalpies ($\Delta H(s)$).

4.4.2 Accuracy of the Methods

In this section the accuracy of the presented methods is discussed. Three significant approximations can be identified as possible sources of error.

The lowest level approximation is related to the GPW method for which an auxiliary Plane-Waves basis is introduce in order to express the fitting densities in the reciprocal space. We have shown that the GPW approach for calculation the MP2 energy introduce negligible error, and the setting specified in Section 4.4.1 are sufficient to provide an error below 10^{-7} a.u per heavy atom [44].

The second approximation refers to the RI methods, for which an auxiliary basis of localized Gaussians is introduced in order to factorize the two electron repulsion integrals. We have tested the accuracy of the RI approximation only in the MP2 energy case, since no reference dRPA energies have been calculated due to the high $O(N^6)$ computational cost. However, it has been shown that the magnitude of the error introduced by RI is of the same order for the MP2 and dRPA energies [10, 43]. The largest deviation observed is related to the case of the Benzene crystal for which the absolute error is 1.78 milli-Hartree corresponding to approximately 10^{-5} Hartree per atom.



(a) **(b)**

Figure 4.5: Accuracy of the numerical quadrature for the RI-dRPA (a) and RI-SOS-MP2 (b) methods with respect to number of quadrature points for three different systems. The reference energy is, in the case of RI-dRPA, the energy obtained employing 100 quadrature points, in the case of RI-SOS-MP2, the exact spin-opposite component of the RI-MP2 energy.

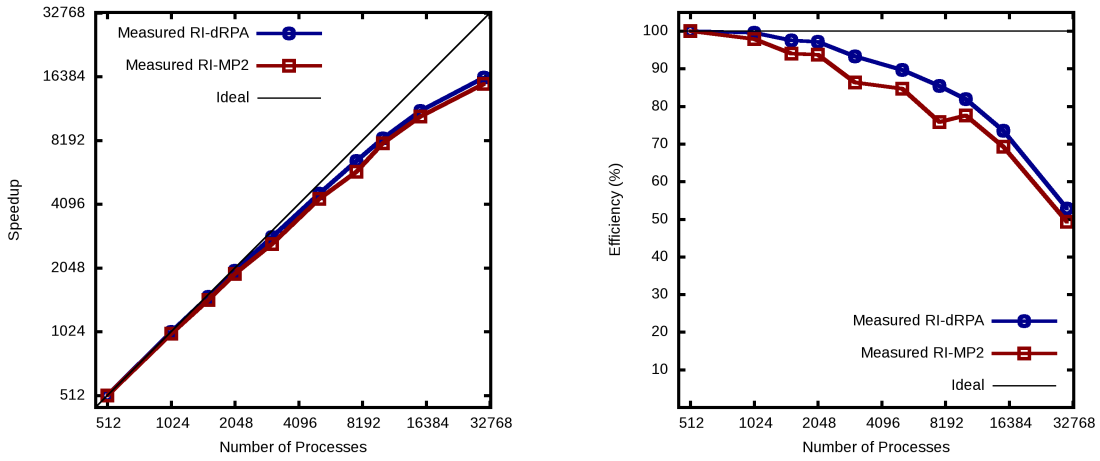
The last possible source of error concerns only the RI-dRPA and RI-SOS-MP2 methods for which an integral quadrature scheme is required. Figure 4.5 shows the convergence of the absolute error with respect to the number of quadrature points for the two methods for three different condensed phase systems. In both cases, an exponential convergence is observed. For the Clenshaw-Curtis quadrature scheme, employed in the dRPA case, $20 - 30$ quadrature points are enough for micro Hartree accuracy. For the minimax approximation, used in the Laplace transformed SOS-MP2 method, the same accuracy is obtained with only $6 - 8$ points.

4.4.3 Performance of the Methods

To assess the performance of the parallel algorithms a system made of 64 bulk water molecules has been chosen. The employed basis set is the cc-TZVP resulting in 265 occupied orbitals, 3648 primary and 8704 auxiliary basis functions. For the RI-dRPA case, 60 integration points have been used for the numerical quadrature.

The speed-up and the parallel efficiency for the RI-MP2 and RI-dRPA methods are reported in Figure 4.6. The Laplace-RI-SOS-MP2 method has not been reported since it is closely related to the RI-dRPA case. Both algorithms show good parallel scalability in a wide range, resulting in an efficiency around 80% for the 10240 processes run. At the full scale-out (30720 processes) the time for computing the RI-MP2 and RI-dRPA energies is 97 and 109 seconds respectively.

The main reason of the reduced efficiency in the RI-MP2 case is load imbalance, since the number of independent ij pairs starts to be of the same order as the number of processes for the largest runs. In the RI-dRPA case, the reduced efficiency is mainly



(a) **(b)**

Figure 4.6: Measured speed up (a) and efficiency (b) with respect to 512 processes for the calculation of the RI-MP2 and RI-dRPA energy of 64 bulk water molecules. RI-dRPA calculation performed employing 60 quadrature points for the numerical integration.

related to the replication of the **B** matrix into each integration group R. Even if this step takes only 32 seconds for the full scale-out run it is significant for the total time.

In Table 4.3 the timing for different benchmark calculations performed with 3200 processes are reported. Together with the time spent for the computation of the correlation energy for the three methods, also the time spent in the first Hartree-Fock SCF cycle and the speed up of the RI-MP2 compared to the canonical MP2 are shown. The time for the first SCF cycle is strongly dominated by the calculation of the four-center electron repulsion integrals over atomic orbitals, this time is reported only to give an idea of the relative computational effort between the SCF procedure and the correlation energy calculation.

The calculation of the RI-SOS-MP2 energy is the fastest in all cases due to the combination of the lower $O(N^4)$ scaling and the relatively small number of quadrature points necessary. Moreover, in this case, the timing are of the same order as the first SCF step.

The RI-dRPA method is as expensive as the RI-MP2 method for the smaller systems, but becomes cheaper than the latter for larger systems due to the better asymptotic scaling of $O(N^4)$.

Finally the RI-MP2 method shows a speed up compared to the canonical MP2 energy calculation that ranges from 9.8 up to a maximum of 17.3.

4.4.4 Performance of the Hybrid CPU/GPU Implementations

In order to assess the performance of the hybrid CPU/GPU implementations the same benchmark systems reported in Table 4.3 for the standard algorithms have been used. The results are shown in Table 4.4, where the data refer to the total execution time

	A	n	o	N_a	t^{RI}_{MP2}	t^{RI}_{dRPA}	$t^{RI}_{\text{SOS-MP2}}$	$t^{\text{SCF}}_{1^{st}\text{cycle}}$	$\frac{t_{\text{MP2}}}{t^{RI}_{\text{MP2}}}$
U	128	2752	192	6784	2.4	2.5	1.3	2.1	12.9
D	128	2992	192	7520	3.3	3.5	1.9	4.9	14.3
FA	120	2760	216	6912	2.8	2.8	1.5	2.0	12.1
64 H_2O	192	3648	256	8704	7.5	6.4	3.4	1.0	17.3
B	192	4128	240	10176	10.4	8.8	4.2	5.5	12.7
PD	144	3936	312	10208	13.7	10.3	4.5	6.9	10.4
SA	176	4144	304	10432	14.6	11.1	4.9	5.0	10.5
CT	168	4152	336	10560	17.2	12.3	5.2	4.5	9.8

Table 4.3: Benchmark calculations for the RI-MP2, RI-dRPA and RI-SOS-MP2 methods, time in min.@CRAY-XK7, 3200 cores. U = Urea, B = Benzene, FA = Formic Acid, SA = Succinic Anhydride, D = 2,3-Diazanaphthalene, PD = Pyromellitic Dianhydride, CT = Cyclotrimethylene-Trinitramine, 64 H_2O = 64 bulk water molecules. A, n, o and N_a represent the number of atoms, basis functions, occupied orbitals and auxiliary basis functions respectively. t^{RI}_{MP2}, t^{RI}_{dRPA}, $t^{RI}_{\text{SOS-MP2}}$ and $t^{\text{SCF}}_{1^{st}\text{cycle}}$ are the time necessary for the RI-MP2, RI-dRPA, RI-SOS-MP2 energy calculations and for the first iteration in the Hartree-Fock self consistent field (SCF) procedure. $\frac{t_{\text{MP2}}}{t^{RI}_{\text{MP2}}}$ is the speedup of the RI method compared to the standard canonical MP2 energy algorithm. All RI-dRPA and RI-SOS-MP2 calculations were performed employing 40 and 10 quadrature points respectively for the numerical integration.

for computing the correlation energies and the time related only to the matrix-matrix multiplication step, labeled for both RI-MP2 and RI-dRPA as "mul". The superscript "H" denotes the case of the hybrid implementations, while the absence of superscript refers to the standard algorithms. The total execution timings in the case of the standard implementations can be found in Table 4.3.

Comparing the total timings for computing the correlation energies with t_{mul}, for the standard algorithms, it can be noted that this step requires between 50% to 80% of the total time. This sets the boundaries for the impact of the acceleration of this task on the speed-up of the overall calculation.

In Table 4.4 the speed-up of the matrix multiplication step is labeled as $\left(\frac{t_{\text{mul}}}{t^{\text{H}}_{\text{mul}}}\right)$. In the case of RI-MP2 the speed-ups range from 3.0 up to 4.3, while in the case of RI-dRPA the results are better resulting in speed-ups that are typically greater than 4.

The acceleration of the matrix-matrix multiplication by GPUs has a significant impact on the overall timing, giving a speed-up of 1.6 for the smaller systems to more than 2 for the larger ones where the matrix multiplication step is clearly dominant. In the limit of very large systems, the speed-up can be expected to be even greater than 3.

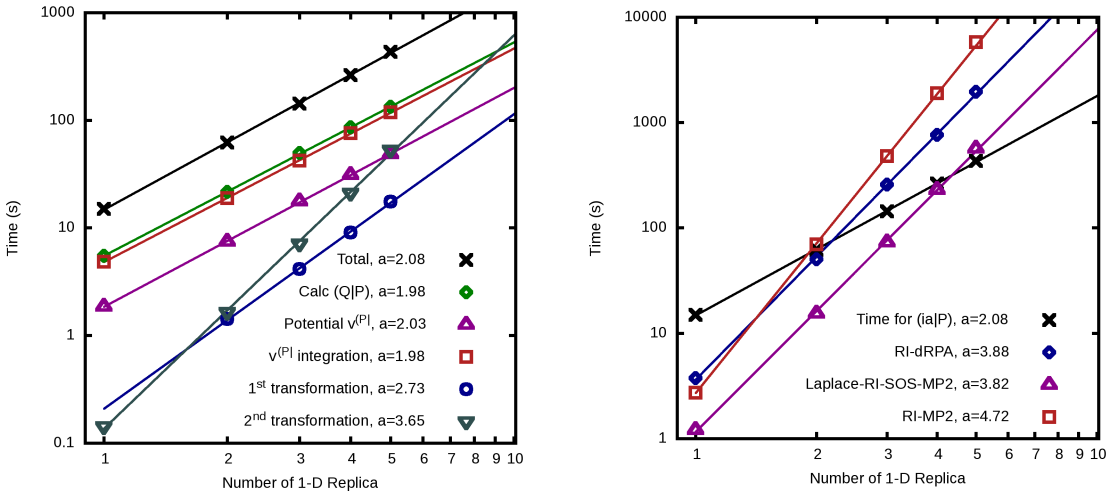
	RI-MP2					RI-dRPA				
	t_{mul}	$t^{\text{H}}_{\text{mul}}$	$\frac{t_{\text{mul}}}{t^{\text{H}}_{\text{mul}}}$	$t^{\text{H}}_{\text{tot}}$	$\frac{t_{\text{tot}}}{t^{\text{H}}_{\text{tot}}}$	t_{mul}	$t^{\text{H}}_{\text{mul}}$	$\frac{t_{\text{mul}}}{t^{\text{H}}_{\text{mul}}}$	$t^{\text{H}}_{\text{tot}}$	$\frac{t_{\text{tot}}}{t^{\text{H}}_{\text{tot}}}$
U	1.37	0.40	3.4	1.49	1.6	1.39	0.32	4.3	1.39	1.8
D	1.80	0.61	3.0	2.10	1.6	1.87	0.44	4.2	2.01	1.7
FA	1.72	0.58	3.0	1.70	1.6	1.61	0.40	4.0	1.58	1.8
64 H_2O	5.00	1.47	3.4	4.19	1.8	4.03	0.91	4.4	3.20	2.0
B	7.14	1.93	3.7	5.16	2.0	5.86	1.36	4.3	4.20	2.1
PD	10.5	2.90	3.6	6.20	2.2	7.46	1.79	4.2	4.65	2.2
SA	11.2	2.62	4.3	6.14	2.4	7.93	1.84	4.3	4.95	2.2
CT	13.7	3.78	3.6	7.44	2.3	9.03	2.01	4.5	5.24	2.3

Table 4.4: Benchmark calculations for the hybrid CPU/GPU implementation of the RI-MP2 and RI-dRPA methods, time in min.@CRAY-XK7, 3200 cores, 200 GPU. The meaning of the row entries is given in Table 4.3. In the table the superscript H refers to the hybrid CPU/GPU implementations, if no superscript is specified the timing are referred to the standard (only CPU) implementation. In the case of RI-MP2 the subscript "mul" refers to the $O(N^5)$ step of the algorithm, that is the $(ia|jb)_{RI}$ integrals generation (Equation (4.5)). In the case of RI-dRPA the subscript "mul" refers to the $O(N^4)$ step in the algorithm, that is the calculation of $\mathbf{Q}(\omega) = 2\mathbf{B}^{\mathbf{T}}\mathbf{B}'(\omega)$ performed as a parallel matrix-matrix multiplication.

4.4.5 System Size Scaling

In order to validate the performance models listed in Table 4.1 and Table 4.2, the time for the individual steps of the algorithms has been measured for systems of increasing size. The test system is based on a supercell containing 32 bulk water molecules, with a cc-DZVP basis set. The supercell has been replicated up to five times in one dimension, giving a final supercell containing 160 molecules. All calculations have been performed employing 1200 cores without GPUs. The timing are reported separately for the calculation of the $(ia|P)$ integrals (in common for all methods, Figure 4.7a) and for the matrix-multiplication part of the correlation energy calculation (different for all methods, Figure 4.7b). The obtained timing have been fitted with the function $y = bx^a$, yielding the measured scaling exponent a associated with each different step.

In the legend of Figure 4.7a the label "Calc (Q—P)" includes all the steps necessary to calculate the two center integrals over the auxiliary basis functions, they are all expected to scale roughly $O(N^2)$. For the calculation of the final $(ia|P)$ integrals more details have been reported separating the calculation of the potential, the integration of the potential to give the $(\mu\nu|P)$ integrals, and the index transformations leading to the final required form $(ia|P)$. As shown by the measured values of a the designed models are confirmed giving an $O(N^2)$ scaling for both the potential generation and integration, and resulting in $a \simeq 3$ and $a \simeq 4$ for the first and second index transformation respectively. It is important to note here the Cholesky decomposition and the triangular inversion steps



(a) **(b)**

Figure 4.7: Shown is the time spent in the various significant part of the energy calculation for the three methods, as a function of the number of replicas of the supercell, containing 32, 64, 96, 128 and 160 molecules of H_2O respectively. The number of quadrature points employed is 20 and 6 respectively for the RI-dRPA and RI-SOS-MP2 methods. (a) reports the timing for the $(ia|P)$ integral generation, that is in common for all methods (see Table 4.1) (b) reports the total time for the $(ia|P)$ integral generation (black line) together with the timing for the most expensive step for each individual method (see Table 4.2). Lines represent a linear two-parameter fit of the form $y = bx^a$. The values of a for each task are reported in the legend.

are not reported since, even having an expected scaling of $O(N^3)$, the time associated with these steps is insignificant compare to the total for all sizes tested.

In Figure 4.7b the reported timing are related to the calculation of the $(ia|jb)_{RI}$ integrals, the $\mathbf{Q}(\omega) = 2\mathbf{B^T B}'(\omega)$ and the $\overline{\mathbf{Q}} = \mathbf{B^T B}'$, respectively for the RI-MP2, RI-dRPA and RI-SOS-MP2 methods. Also in this case the formal scaling of the individual step is confirmed to be $\simeq 5$ for RI-MP2 and $\simeq 4$ for the other two methods. The smaller prefactor observed for RI-SOS-MP2 compared to RI-dRPA is due to the lower number of points required in order to reach the same integration accuracy in the numerical quadrature.

4.4.6 Solid LiH

Due to its favorable features, such as large band gap, simple unit cell and absence of heavy atoms, the LiH crystal has been widely used as benchmark system in condensed phase electronic structure calculations [44, 50, 106, 116–118, 134]. In this section the estimated complete basis set (CBS) limit of the RI-MP2 and RI-dRPA contributions to the cohesive energy of LiH crystal are reported and discussed. In order to do so, two extrapolations have been performed, the first with respect to the cell size going to infinity, the second with respect to the basis set.

The counterpoise corrected contributions to the cohesive energy of LiH for various basis set and cell sizes together with the extrapolated values are summarized in Table 4.5.

Table 4.5: All-electron pseudopotential HF@PBE, RI-dRPA@PBE and RI-MP2 contributions to the counterpoise corrected cohesive energies in mE_h of LiH at the experimental geometry ($a = 4.084$ Å) for various basis set and cell sizes. The text discusses how the extrapolated numbers (italic type) have been obtained.

	$2\times2\times2$	$3\times3\times3$	$4\times4\times4$	Extr. ($E_X^{n\to\infty}$)
		HF@PBE		
cc-DZVP	-131.91	-134.80	-135.00	
cc-TZVP	-124.84	-128.07	-128.31	
cc-QZVP	-124.41	-127.63	-127.88	
		RI-dRPA@PBE		
cc-DZVP	-27.05	-28.53	-28.95	*-29.19*
cc-TZVP	-38.42	-40.11	-40.62	*-40.89*
cc-QZVP	-41.86	-43.73	*-44.28*	*-44.59*
Extr. ($E_{X\to\infty}^{n\to\infty}$)				*-46.42*
		RI-MP2		
cc-DZVP	-29.25	-30.30	-30.57	*-30.75*
cc-TZVP	-38.00	-39.33	-39.68	*-39.91*
cc-QZVP	-40.57	-41.99	*-42.36*	*-42.60*
Extr. ($E_{X\to\infty}^{n\to\infty}$)				*-44.10*

The calculation of the RI-dRPA energies as well as the EXX contributions have been performed with input electronic orbitals and orbital energies obtained from self-consistent PBE calculation. They are denoted respectively as RI-dRPA@PBE and HF@PBE. The calculation have been performed up to the supercell $4\times4\times4$ for the cc-DZVP and cc-TZVP basis and up to supercell $3\times3\times3$ for the cc-QZVP. In the following discussion, the energy contribution for a given pair of cell size/basis set is labeled as E_X^n, with X denoting the basis (D, T, Q, ∞) and n the number of repeated unit cells $(1, 2, 3, \infty)$

The estimate for the size converged limit, for a given basis set ($E_X^{n\to\infty}$), has been obtained employing the extrapolation formula $E_X^n = E_X^{n\to\infty} + S(n\times a)^{-3}$ where n is the number of repetitive cells considered and a is the lattice parameter. The choice of the exponent -3 used in the extrapolation has been inspired by the long range behavior of both MP2 and dRPA pair energy, following the London law C_6/d_{ij}^6, with d_{ij} being the distance between the center of two charge distributions [59, 172], and integrating over all pairs in the crystal for which $d_{ij} \geq d$. For all the basis sets considered, the $E_X^{n\to\infty}$ value has been obtained by a three points extrapolation for cell sizes ranging from $2\times2\times2$ to $4\times4\times4$. In the case of the cc-QZVP basis the additional point E_Q^4 has been obtained as $E_Q^3 \times \dfrac{E_T^4}{E_T^3}$. This extrapolation is justified by the observation that the ratio between the energies per formula unit for two consecutive cell sizes $\left(\frac{E_X^n}{E_X^{(n+1)}}\right)$ converges quickly with respect to the basis set for both MP2 and dRPA. In fact, as shown in Table 4.6, the difference of $\frac{E_X^2}{E_X^3}$ in going from the cc-TZVP to the cc-QZVP basis is of the order of 10^{-4}

Table 4.6: Ratio between the energies per formula unit of bulk LiH for two consecutive cell size $\left(\frac{E_X^n}{E_X^{(n+1)}}\right)$. X=D,T,Q for the cc-DZVP, cc-TZVP and cc-QZVP respectively.

	RI-MP2		RI-dRPA@PBE	
	$\frac{E_X^2}{E_X^3}$	$\frac{E_X^3}{E_X^4}$	$\frac{E_X^2}{E_X^3}$	$\frac{E_X^3}{E_X^4}$
cc-DZVP	0.9794	0.9951	0.9865	0.9961
cc-TZVP	0.9813	0.9954	0.9884	0.9965
cc-QZVP	0.9815		0.9881	

for both RI-MP2 and RI-dRPA. Moreover, the difference between the computed result E_Q^3 and the extrapolated $E_Q^2 \times \frac{E_T^3}{E_T^2}$ is of the order of few tenth of μE_h for both RI-MP2 and RI-dRPA, validating the approach.

In order to obtain the CBS, the size extrapolated results for each basis set have been finally extrapolated with the cubic interpolation formula $E_X^{n\to\infty} = E_{X\to\infty}^{n\to\infty} + AX^{-3}$ ($X = 2, 3, 4$ for cc-DZVP, cc-TZVP and cc-QZVP respectively) [120, 133]. The fits are shown in Figure 4.8 and the extrapolated values reported in Table 4.5.

Finally, the complete basis set limit of the cohesive energy of LiH is estimated to be 44.10 and 46.42 mE_h for RI-MP2 and RI-dRPA respectively. The obtained result in the RI-MP2 case is in excellent agreement with our previously reported MP2 value of 44.09, showing that the error introduced by the RI approximation is negligibly small. Considering the converged HF@PBE and HF values to be 127.9 and 132.0 mE_h respectively, it can be observed that, while the obtained value for the dRPA@PBE contribution to the cohesive energy is slightly larger that the pure MP2 result, the total cohesive energy given as (HF+dRPA)@PBE is underbinding compared to the HF+MP2 value. This behavior is in agreement with the trend observed for the atomization energies reported in several previous studies [9, 10, 12, 138].

4.4.7 Cohesive Energy of Molecular Crystals

The counterpoise corrected cohesive energy for several molecular crystals, calculated at different level of theory using the experimental crystal cell parameters and employing the cc-TZVP basis, are reported in Table 4.7.

Together with the RI-MP2 results also the MP2 values, taken from ref. [44], have been reported as reference in order to assess the accuracy of the RI approximation. As shown in Table 4.7, the error introduce by the RI approximation is, also in this case, negligibly small, resulting in an error of a few tenth of kJ/mol in the final value of cohesive energy.

In the case of EXX/dRPA method, the effect of different input orbitals in the computation of the counterpoise corrected cohesive energies has been tested. The tested reference wave-functions have been chosen with a decreasing fraction of non-local exchange, ranging from the pure Hartree-Fock (100%), to PBE (0%), passing through B3LYP [110, 112, 113]

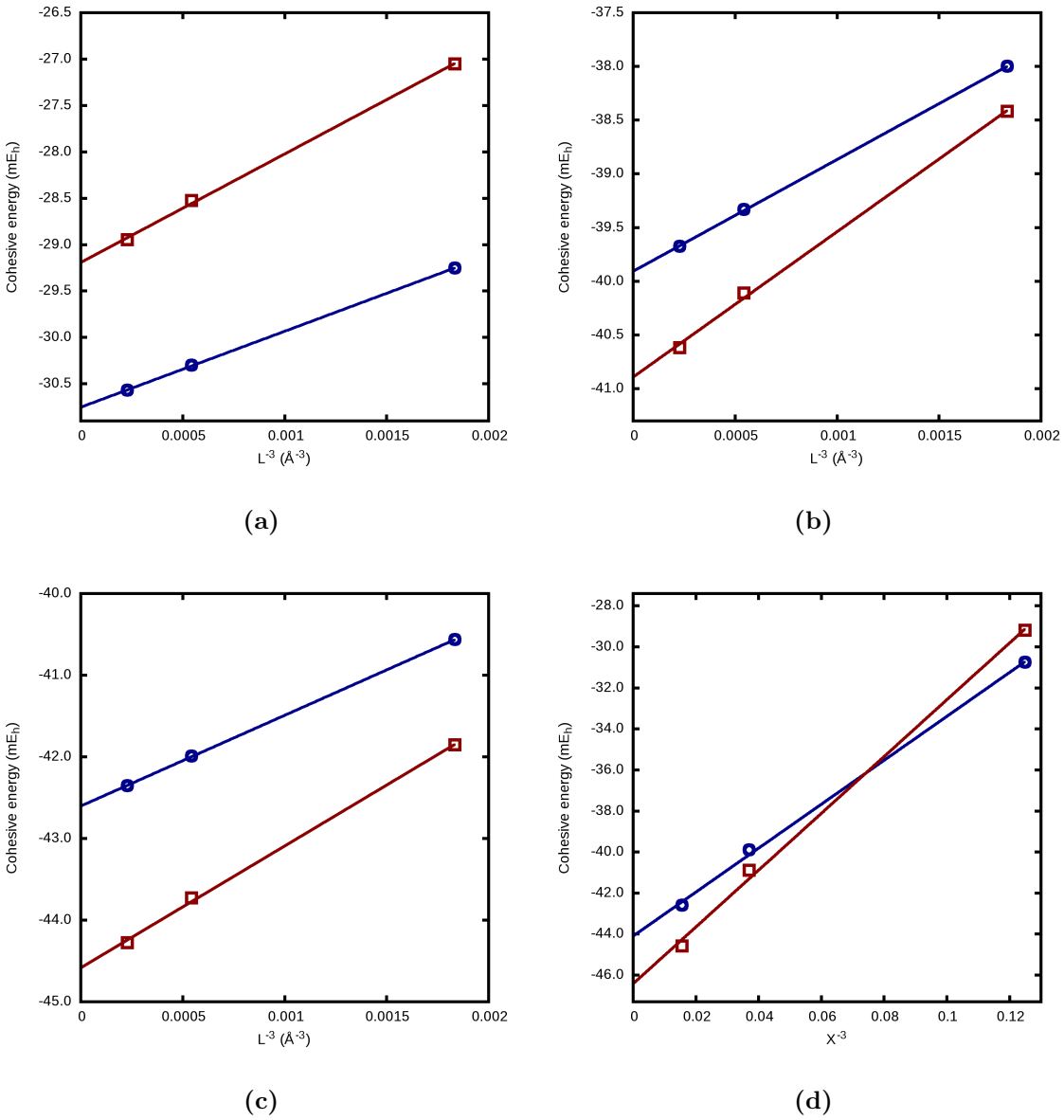


Figure 4.8: Convergence of the all-electron pseudopotential RI-dRPA@PBE (red squares) and RI-MP2 (blue circles) contribution to the cohesive energy of LiH at the experimental geometry ($a = 4.084$ Å) with respect to the cell size and basis set. L is the length of the cell edge ($L = n \times a$) and $X = 2, 3$ and 4 for cc-DZVP, cc-TZVP and cc-QZVP respectively. (a) Extrapolation with respect to the cell size for the cc-DZVP basis set. (b) Extrapolation with respect to the cell size for the cc-TZVP basis set. (c) Extrapolation with respect to the cell size for the cc-QZVP basis set. (d) Extrapolation with respect to the basis set for the size-converged cohesive energies.

(20%) and PBE0 [173] (25%). As shown in Table 4.7 the decreasing fraction of non-local exchange, in general, results in an increase of the dRPA and a decrease of the EXX contributions to the cohesive energy. As noted in previous studies on vdW bonded molecular crystals [168, 169], these two effects have roughly the same magnitude resulting, in most cases, in a cancellation in the final value of the EXX/dRPA cohesive energy. The smallest root-mean-square deviation (RMSD), compared to the experimental sublimation

Table 4.7: Counterpoise corrected cohesive energy ($-E_{coh}^{CP}$) in kJ/mol for the the molecular crystals of B = Benzene, FA = Formic Acid, SA = Succinic Anhydride, D = 2,3-Diazanaphthalene, PD = Pyromellitic Dianhydride, U = Urea, CT = Cyclotrimethylene-Trinitramine. The sign of the E_{coh}^{CP} has been changed in order to be compared with the experimental sublimation enthalpies $\Delta H(s)$. $\Delta H(s)$ have been taken from the supporting information of Ref [111], see also http://webbook.nist.gov/chemistry/.

	B	FA	SA	D	PD	U	CT	RMSD
Exp $\Delta H(s)$	45	68	81	83	83	92	112	
HF	-21.2	26.3	38.6	-5.7	31.3	55.8	49.8	58.1
MP2	58.8	55.5	81.2	79.7	123.4	94.6	113.7	16.9
RI-MP2	58.8	55.6	81.1	79.8	123.2	94.6	114.0	16.8
(HF+dRPA)@HF	29.2	45.0	68.1	45.7	92.0	81.1	91.1	20.6
dRPA@HF	50.3	18.6	29.5	51.4	60.7	25.3	41.3	
(HF+dRPA)@PBE0	36.3	47.0	69.9	49.2	94.1	83.4	83.9	19.9
dRPA@PBE0	60.6	27.7	36.1	54.6	71.3	33.2	36.0	
HF@PBE0	-24.3	19.3	33.8	-5.3	22.8	50.2	47.8	
(HF+dRPA)@B3LYP	37.1	46.9	69.9	49.1	94.1	83.4	82.8	20.1
dRPA@B3LYP	62.4	26.9	34.7	54.7	70.3	32.8	32.5	
HF@B3LYP	-25.4	20.0	35.2	-5.6	23.9	50.6	50.3	
(HF+dRPA)@PBE	37.6	45.0	67.3	44.8	91.0	81.1	75.0	23.2
dRPA@PBE	65.9	30.1	37.2	55.1	72.8	35.5	26.8	
HF@PBE	-28.3	14.8	30.2	-10.3	18.2	45.7	48.2	

enthalpies, is observed for the case in which PBE0 is used, even if the difference in RMSD with the other EXX/dRPA cases is quite small.

Finally, the obtained cohesive energies with the EXX/dRPA method are systematically underbinding the molecular crystals compared to the MP2 case, giving a slightly larger RMSD.

4.5 Conclusions

With the present chapter, a novel method for the resolution of identity approximation applied to the calculation of electron repulsion integrals over molecular orbitals, based on the Gaussian and Plane Waves approach is introduced. The ground foundation of this method lies in the way the three center electron repulsion integrals of the type $(\mu\nu|P)$ are computed, that is, by direct integration between the product of the Gaussian basis function

$(\mu\nu|$ and the electrostatic potential arising from the RI fitting densities $(P| = \sum_R (R|L^{-1}_{PR}$. This approach has been shown to be efficient, accurate and robust for periodic systems, furthermore it displays a measured scaling of the computational effort that grows only quadratically with the system size. In addition to that, it offers a straightforward way for parallel implementation.

The RI approximation has been applied to the calculation of the correlation energy at the MP2, SOS-MP2 and dRPA level of theory for finite and extended systems. Massively parallel algorithms have been developed for each of these methods, displaying excellent parallel scalability and efficiency up to ten thousands of processes. Furthermore, it has been shown that a hybrid CPU/GPU implementation can result in speed-ups to individual steps of the algorithm up to a factor 4.5 and around a factor 2 for the global calculation compared to the standard only CPU implementation.

The RI approximation introduces negligible error in the final energy evaluation while giving a speed-up, in the case of RI-MP2, of a factor 10 to 15 compared to the standard GPW-MP2 implementation. In the SOS-MP2 and dRPA cases, the energy evaluation is carried out by a numerical integration procedure that allows to reduce the computational effort to $O(N^4)$. The numerical quadrature schemes has been shown to converge exponentially with respect to the number of grid points, such that, for systems with a sizable gap, 6-8 and 20-30 points are enough for micro Hartree accuracy in the case of SOS-MP2 and dRPA respectively.

Several benchmark calculations have been reported, showing that correlation energy calculations, at the different level of theory presented, can be performed within minutes for systems containing hundreds of atoms and thousands of basis functions.

Chapter 5

Enabling Simulation at the Fifth Rung of DFT: Large Scale RPA Calculations with Excellent Time to Solution [174]

The Random Phase Approximation (RPA), which represents the fifth rung of accuracy in Density Functional Theory (DFT), is made practical for large systems. Energies of condensed phase systems containing 1000s of explicitly correlated electrons and 1500 atoms can now be computed in minutes and less than one hour, respectively. GPU acceleration is employed for dense and sparse linear algebra, while communication is minimized by a judicious data layout. The performance of the algorithms, implemented in the widely used CP2K simulation package, has been investigated on hybrid Cray XC30 and XK7 architectures, up to 16384 nodes. Our results emphasize the importance of good network performance, in addition to the availability of GPUs and generous on node memory. A new level of predictivity has thus become available for routine application in Monte Carlo and molecular dynamics simulations.

5.1 Introduction

Density Functional Theory (DFT) is a Nobel prize winning theory that has become the most widely used quantum mechanical tool in chemistry and physics for predicting properties of materials ranging from single molecules to condensed phase systems. Formally, DFT is exactly equivalent to the Schrödinger equation [1, 2]. Practical DFT calculations replace the unknown exact exchange and correlation energy by one of the various models (functionals) of electron correlation. The success of DFT can be attributed to the excellent 'price-performance ratio' that simple semi-local models of electron correlation (GGAs) offer. GGA DFT requires only the electron density and its gradient as an input to the functional, and this simple form leads to a low computational cost, scaling moderately ($O(N^3)$) with system size (N). Systems containing several hundreds of atoms, necessary and sufficient to describe a wide of range of condensed phase phenomena of scientific and technological impact, can now be routinely studied. To optimize atomic structures or to perform *ab*

initio molecular dynamics (MD) and Monte Carlo (MC), time-to-solution is essential, as 10s to 10'000s of sequential energy/force calculations are required. Consequently, only if the time of a single energy/force calculation drops below a few minutes, *ab initio* MD and MC become practical. Enabling MD and MC has been a prime motivation for massively parallel implementations of GGA DFT in the past, and is, at a more advanced level, also a target for the current work.

To make progress on the way to high accuracy and thus to the predictivity of the theory, we focus on correlation models beyond GGA DFT. To illustrate the need for functionals beyond GGA, the failure of standard GGA functionals to yield van der Waals or dispersive interactions can be used as one example. The absence of these important weak interactions leads to gross errors and qualitatively wrong predictions of seemingly simple properties. For example, the density of water is underestimated by roughly 20% by the GGA BLYP functional[175–177], and GGA ice sinks in, rather than floats on, GGA water. More advanced correlations models take ingredients beyond the electron density into account, and utilize this additional data to improve the quality of the results. The class of functionals that include the virtual orbitals, the so called fifth rung of DFT, is an essential step forward. Indeed, for one such model (MP2 theory), we have recently demonstrated by means of MC simulations that it yields a computed density of water within 2% of experiment, a 10-fold improvement over BLYP [177]. This improvement results in a correct prediction of the most famous anomaly of water, namely that ice floats.

We first clarify the nomenclature and properties of 'fifth rung' functionals, before we discuss the computational aspects. Perdew *et al.*[3] classified the known functionals on a "Jacob's ladder", leading from a 'Hartree-world' to a metaphorical heaven of chemical accuracy. Each rung of the ladder introduces more descriptors of the electronic system and yields models with improved accuracy. This chapter deals with the most advanced, fifth rung, which include 1) the density, 2) its gradient, 3) the kinetic energy density, 4) the occupied, and most importantly 5) the virtual orbitals. By capturing physical phenomena not accounted for in semi-local DFT, the fifth rung represents essential progress, including best performance in various benchmark sets [10, 18, 19, 178]. The various functionals on the fifth rung include virtual orbitals either via the random phase approximation (RPA) correlation [4–12] or via terms derived from second-order Møller-Plesset (MP2) theory, in the form of double hybrids (DH) [13–19]. The focus of this manuscript is on RPA, which has favorable properties both from a computational and physical point of view. Nevertheless, results from MP2 and DH approaches are discussed as well, given the similarity of the methods and their identical location on "Jacob's ladder".

It is not surprising that the benefits of improved accuracy come with increased computational requirements. Unfortunately, this increase from the commonly used second rung to the fifth rung is so large that until our recent work[177] no condensed phase MD or MC had ever been performed. Indeed, the formal computational effort of RPA and MP2 calculations, in their canonical formulation, scales as $O(N^6)$ and $O(N^5)$ respectively, compared to the $O(N^3)$ of the second rung. An aggravating fact is that fifth rung methods need significantly more basis functions, the proper measure of system size N, than the lower rungs to deliver converged results [133, 134]. Prefactors are such that already for a few dozen atoms second and fifth rung approaches differ by roughly three orders of magnitude in computational demand.

Here, we present recent algorithmic and computational progress that mitigates these

requirements and allows for RPA calculations with excellent time to solution on massively parallel hybrid architectures. As we will discuss in detail below, this is based on recent developments:

- An efficient formulation of RPA that exhibits a manageable $O(N^4)$ scaling with small prefactor [43, 179].
- A novel algorithm, named RI-GPW, for the computation of the needed integrals, which includes a data layout and communication schemes that result in a massively parallel implementation [131].
- A new implementation that brings all $O(N^4)$ steps to the GPU, including dense and sparse linear algebra [180].

As a result, using the latest hybrid XC30 computer architecture, energy calculations on typical simulation cells of liquid water (64 H_2O), reach the one minute per step limit already at 512 nodes. This makes MC simulations on this important system now routinely feasible, and represents a breakthrough in the field. The energy of 1000s of correlated electrons (256 H_2O) can be computed in less than ten minutes on 4096 XC30 nodes, and a calculation on 1500 atoms (512 H_2O) is feasible in less than one hour on 16384 XK7 nodes. We therefore anticipate that this level of simulation will be used with increasing frequency to tackle further important open questions with relevance to chemistry, physics, biology, energy and materials science.

5.2 Current state of the art

In order to perform these calculations with acceptable time to solution, and thus to extend the applicability of these methods to larger systems, the development of efficient algorithms and scalable implementations is of prime interest and an active topic of research in the community. It is now well established that the resolution of identity (RI) is a powerful technique to reduce both the prefactor and the scaling of RPA and MP2 calculations [43, 46, 47, 68–73, 127, 137, 138]. As discussed below, within direct RPA, the RI approach speeds-up the calculation of the integrals, reduces the memory requirements, and additionally also allows for reducing the scaling from $O(N^6)$ to $O(N^4)$ [34, 36, 43]. This significant reduction in scaling is key for the application to large systems.

Several RPA implementations have been reported in the literature, formulated in term of Gaussian [34, 36, 43], numeric [12, 138] and plane-wave (PW) [7, 157, 167, 168, 181] basis sets. Gaussian basis functions, known for their efficiency for gas phase calculations, but also used by us for the condensed phase, have been used by Eshuis and coworkers [43], reporting as the largest test case the calculation of the RPA correlation energy of the octapeptide Angiotensin II (146 atoms, 1117 atomic orbitals) requiring 18.9 h on a Xeon X7550 2.00 GHz CPU. PW basis sets are most commonly used for condensed phase calculations and RPA has been successfully employed for studying solids [158–162, 182], surfaces [163–166, 178] and van der Waals crystals [167–169]. Also in this basis, the computational cost limits the system size. In a very recent paper that studies low scaling RPA methods, current applications are reported to be limited to $O(100)$ electrons [179]. Although no information about the actual time to solution is reported, Harl *et al.* [160]

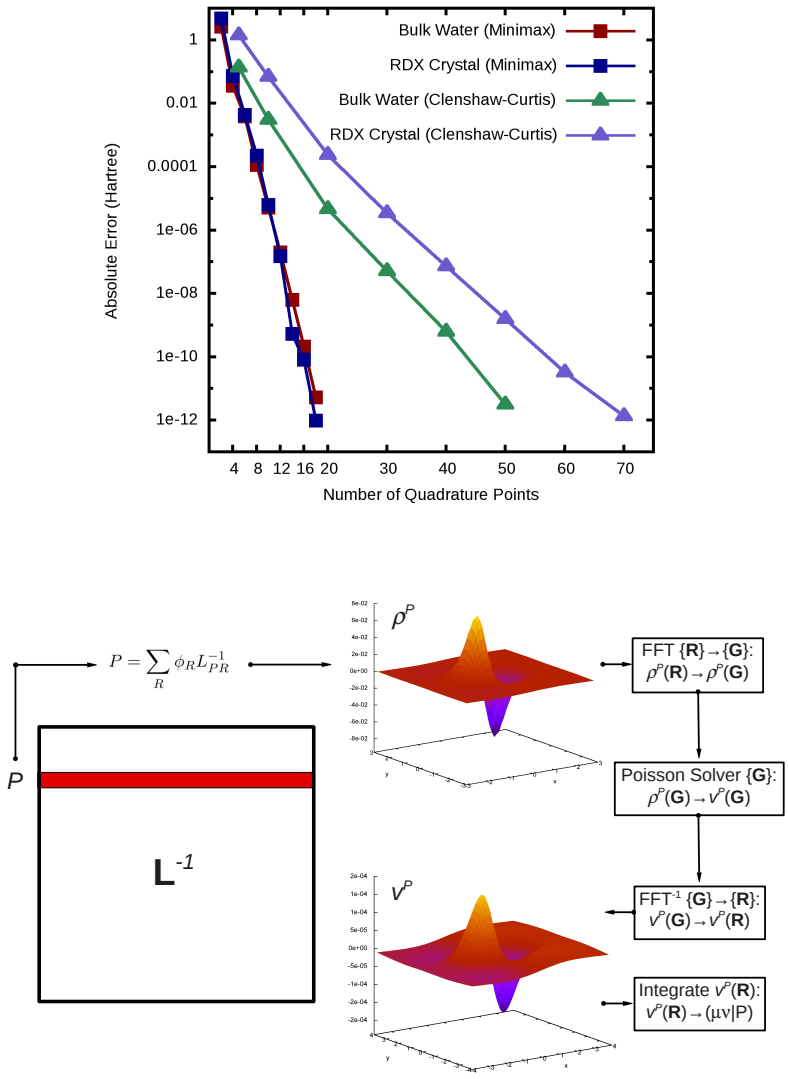


Figure 5.1: Top: Shown is the exponential convergence of the energy expression Equation (5.2) with respect to the number of integration points for two different quadrature schemes discussed in Section 5.3.2 (RDX = Cyclotrimethylene-Trinitramine). This greatly contributes to the efficiency of the $O(N^4)$ RI-GPW RPA scheme. Bottom: Characteristic for the RI-GPW RPA approach is the calculation of the matrix of three-center ERI $(\mu\nu|P)$ using a FFT based Poisson solver, yielding for a give P the sparse matrix in linear scaling time. For each $|P)$, calculations can be performed independently, contributing to excellent scalability.

mentioned that their implementation scales to 100 CPU cores for the $O(N^4)$ computational task. In a recent RPA study on water ice, Macher and coworkers[183] employed 28 H_2O molecules for their largest unit cell.

Contrary to the RPA case, several parallel MP2 and RI-MP2 implementations have been discussed in detail in the past [75–85, 139, 184, 185]. These methods are briefly discussed here, as they share with RI-RPA an algorithmically similar approach to compute the integrals, but ultimately differ due to an $O(N^5)$ scaling. For gas phase molecules,

parallel canonical MP2 calculations have been reported[44, 78, 88], for example requiring 4.8h on 8192 cores for C_{60}. However, it is now clear that RI-MP2 approaches are superior in performance compared to cannonical MP2, without significant loss in accuracy [131]. The largest gas phase RI-MP2 calculation reported, is recent work by Katouda and Nakajima[185], in which, with a new MPI/OpenMP hybrid parallel algorithm, the RI-MP2 energy of interacting nanographene flakes $(C_{150}H_{30})_2$ (9840 basis functions) was calculated in 65 minutes employing 71288 CPU cores of the K computer. When considering systems in the condensed phase, due to the 3D nature and periodic boundary conditions, the calculations become computationally more demanding, and for large systems alternative methods have been developed. In this respect, Maschio presented a parallel local MP2 implementation for periodic systems making use of density fitting [85] reporting a calculation on a metal organic framework (MOF-5) crystalline structure (106 atoms and 2884 atomic orbitals per unit cell) performed in less than 24 h on 53 processors. Recently, we have reported[131] on our parallel RI-MP2 implementation, based on the RI-GPW scheme discussed below. In that work we have shown that our RI-MP2 implementation demonstrates excellent time to solution for systems in the condensed phase containing hundreds of atoms and thousands of basis functions. As an example, for the Cyclotrimethylene-Trinitramine molecular crystal, better known as the RDX explosive (168 atoms and 4152 atomic orbitals per unit cell), the RI-MP2 energy has been calculated in 7.4min on 200 nodes of a Cray XK7 [131]. The efficiency of this scheme is what has enabled our current RI-GPW RPA implementation, and will be discussed in the following.

5.3 Recent Innovations

5.3.1 $O(N^4)$ RI-RPA Theory

In this section, the canonical expression for the direct RPA correlation energy (E_c^{RPA}) is reformulated in such a way that the scaling changes from $O(N^6)$ to $O(N^4)$. This essential step is based on a resolution of identity (RI) technique, recently introduced in RPA by Furche and co-authors [43].

The canonical form of the RPA correlation energy is given by

$$E_c^{\mathrm{RPA}} = \frac{1}{2}\mathrm{Tr}(\mathbf{M}^{1/2} - \mathbf{A})$$

A straightforward computation implies calculation of $\mathbf{M}^{1/2}$. For the 512 H_2O system, $\mathbf{M}$ would be a dense, square, 56M $\times$ 56M matrix, its diagonalization is beyond the capabilities of current supercomputers. The matrix elements of $\mathbf{A}$ and $\mathbf{M}$ are related to the six-dimensional electron repulsion integrals $(ia|jb)$ over canonical molecular orbitals (MO ERIs), and the Kohn-Sham orbital energies ϵ_i and ϵ_a. These ERIs are important quantities in the following, and are indexed by i, j and a, b, which refer to occupied (o) and virtual (v) molecular orbitals (MOs) respectively. The dimension of $\mathbf{M}$ is $(ov) \times (ov)$, so that the number of rows scales as $O(N^2)$, the memory requirements as $O(N^4)$, and the flops as $O(N^6)$, since both o and v are proportional to system size N.

To avoid this steep scaling, an auxiliary Gaussian basis P, Q of size $N_a \propto N$ is

introduced for an RI approximation

$$(ia|jb) \simeq (ia|jb)_{\mathrm{RI}} = \sum_{RQ} (ia|R)(R|Q)^{-1}(Q|jb) = \sum_{P} B_P^{ia} B_P^{jb}. \tag{5.1}$$

The three center matrix $\mathbf{B}$, given by $B_P^{ia} = \sum_R (ia|R) L_{PR}^{-1}$ with L_{PR}^{-1} obtained from the Cholesky decomposition of the two center $(R|Q)$, is thus sufficient to compute the integrals. This $\mathbf{B}$ requires only $O(N^3)$ memory, for the example given above the dimensions are 56M $\times$ 70K, a nearly 1000-fold improvement over $\mathbf{M}$, but nevertheless corresponding to 28 TB of RAM.

Most importantly, exploiting the structure of $\mathbf{M}$ and the factorization of ERIs in terms of $\mathbf{B}$, it can be shown [43] that $E_c^{\text{RI-RPA}}$ is identical to the frequency integral

$$E_c^{\text{RI-RPA}} = \frac{1}{2} \int_{-\infty}^{+\infty} \frac{d\omega}{2\pi} \mathrm{Tr}(\log(\mathbf{1} + \mathbf{Q}(\omega)) - \mathbf{Q}(\omega)). \tag{5.2}$$

Here, the frequency dependent $N_a \times N_a$ matrix $\mathbf{Q}(\omega)$ is given by matrix multiplications as $\mathbf{Q}(\omega) = 2\mathbf{B}^T \mathbf{G}(\omega) \mathbf{B}$, where $\mathbf{G}(\omega)$ is a diagonal matrix that can be computed easily from the Kohn-Sham eigenvalues and frequency. The integral can be computed by numerical quadrature with excellent accuracy using a small number of integration points N_g (independent of system dimension) as shown in Figure 5.1. For a given ω, the computation of the integrand requires $O(N_a N_a ov) = O(N^4)$ operations for the matrix multiplications implied by the definition of $\mathbf{Q}$, and $O(N^3)$ for the evaluation of the trace of the logarithm. The total operation count has thus been reduced from $O(N^6)$ to $O(N^4)$, enabling practical large scale calculations. Within this framework, an $O(N^4)$ formalism for forces have also been established [186].

5.3.2 The RI-GPW RPA method

The RI Gaussian and Plane Waves (RI-GPW) RPA method, recently introduced by us[131], accomplishes the evaluation of Equation (5.2) efficiently and in a highly scalable way. Ref. [131] provides an in-depth description, so here we present the high-level structure, and emphasize recent improvements, implemented to enable the presented calculations. The method consists of two consecutive steps, namely the computation of the matrix elements of $\mathbf{B}$, and the frequency integration of Equation (5.2).

The RI-GPW method

The RI-GPW method deals with the computation of the matrix elements of $\mathbf{B}$, and is equivalent to computing $O(N_a ov)$ 6-dimensional integrals of Coulombic nature. Also this step scales as $O(N^4)$, and is, contrary to the evaluation of Equation (5.2), highly complex in nature and potentially the bottleneck in all but the largest calculations. Referring again to the 512 H_2O system, this step amounts to computing $3.9 \cdot 10^{12}$ 6D integrals, each with non-vanishing contributions from the complete simulation cell.

The parallelization for computing the $(ia|P)$ ERIs within the RI-GPW approach is achieved with a dual-level hybrid OpenMP/MPI scheme, and a careful process layout. To

this end, the N_p processes available in total are split in N_G groups, each consisting of N_w processes ($N_p = N_G N_w$). The first level of parallelization corresponds to distributing the work performed for a single given auxiliary basis function ϕ_P or vector $|P) = \sum_R \phi_R L^{-1}_{PR}$ over the group of N_w processes. Since $N_a > N_p$ it is favorable to make N_w small, leading to excellent efficiency for this step. However, the $O(N^2)$ memory requirements on the first level, are such that the availability of parallelism is essential to run the large systems presented here. The second level of parallelization corresponds to a distribution of the nearly independent calculations on $|P)$ over N_G groups. The fact that these calculations are independent is a major advantage of the RI-GPW technique.

On the first level, the details of the operations are complicated, involving parallel FFTs, halo-exchanges, and sparse matrix multiplications, ultimately computing the electrostatic energy of various Gaussian-based charge distributions. Fortunately, it corresponds to the standard parallelization scheme and modus of operation for GGA DFT calculations in our simulation package CP2K[91]. As such, this step has been streamlined over the past 10 years, including interprocess, multi-threading, and single core optimizations, and will not be discussed in detail, except for the newly developed GPU-accelerated sparse matrix library. As sketched in Figure 5.1, the initial steps consist of a density $\rho(\vec{R})$ calculation on a real space grid, the computation of the associated electrostatic potential $v(\vec{R})$ using a spectral approach, and the final numerical integration of $v(\vec{R})$ over Gaussian basis functions $(\mu\nu)$. Exploiting sparsity, the calculation of the $(\mu\nu|P)$ matrix is $O(N)$ for each given $|P)$. This first step only contributes $O(N^2)$ to the computational workload, but contributes significantly to the total runtime for smaller systems. Its efficiency is key to the success of the RI-GPW RPA method. This step is followed by two integral transformations from the atomic orbital $(\mu\nu|P)$ to molecular orbital $(ia|P)$ basis, performed as two sparse matrix multiplications with comparatively small ($N \times o$ and $N \times v$) matrices. The first index transformation benefits from sparsity, and contributes $O(N^3)$ to the total workload, while the second index transformation is asymptotically dominating with $O(N^4)$ flops. With the new sparse matrix library, both transformations run efficiently on the GPU. At the end of the RI-GPW integral algorithm each process stores part of the matrix elements B^{ia}_P, and the RPA frequency integration can be performed.

RPA frequency integration

For large systems, the total time necessary for the computation of $E_c^{\text{RI-RPA}}$ depends linearly on the number of quadrature points employed in the numerical quadrature of the integral in Equation (5.2). For the calculation of the integral, two different quadrature schemes have been implemented and tested, one is based on the Clenshaw-Curtis (CC) numerical quadrature [152], the other on the Minimax approximation [150, 187].

The CC scheme has been originally introduced for the calculation of $E_c^{\text{RI-RPA}}$ by Eshuis *et al.* [43] and is based on the mapping $\omega = a\cot(t)$ to the interval $t \in [0, \pi]$, a being a scaling parameter optimized according to a diagonal approximation of the RPA excitation energies

The Minimax approximation has been successfully applied for the evaluation of the spin-opposite MP2 energy [149] within the Laplace Transformation method [62, 63] and has been introduce only recently for the calculation of the RPA energy by Kaltak and coworkers [179]. As reported in detail in their work, given N_q quadrature points, the MM

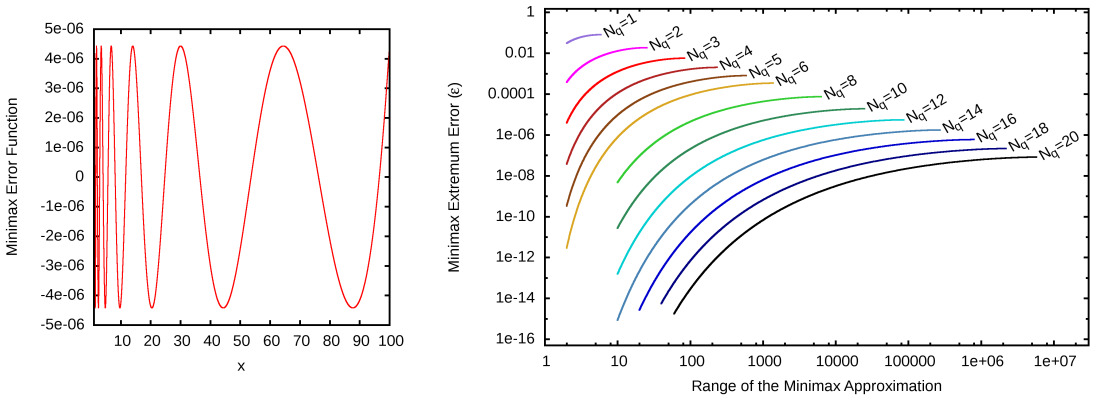


Figure 5.2: Left: minimax error function $\eta(x)$ for $R_c = 100$ and 8 quadrature points. Right: absolute maximum and minimum error (ϵ) of $\eta(x)$ as a function of R_c for different number of quadrature points.

weights $\{w_i\}$ and abscissa $\{\omega_i\}$ are obtained by optimizing the error function

$$\eta(x) = \frac{1}{x} - \frac{1}{\pi}\sum_{i}^{N_q} w_i \left[\frac{2x}{x^2+\omega_i^2}\right]^2 \tag{5.3}$$

such that it creates an alternant in the scaled interval $[1, R_c]$ over which the integration is carried out. More specifically, within the Minimax approximation, the error function satisfies

$$\eta(x_i) = (-1)^i \epsilon, \quad \forall i = 1, 2, \ldots 2N_q + 1 \tag{5.4}$$

for which the $\{x_i\}$ represent the positions of the local maximum and minimum of $\eta(x)$, while ϵ represent the error extremum. In this way $\eta(x)$ guarantees for the best approximation in the given interval due to the alternation of minimum and maximum errors, equal in absolute value and opposite in sign. Figure 5.2 (left) shows the resulting $\eta(x)$ for $N_q = 8$ and $R_c = 100$.

There are two other appealing features of the MM scheme. The first is that the range of the approximation R_c, contrary to the CC case, can be obtained solely from the bare orbital energies and thus is calculated prior to the calculation of the correlation energy. The second is related to the property that, for a given number of quadrature points N_q, an upper bound exists $R_c^{\max}$ such that $\eta(x)$ is strictly decaying for $x > R_c^{\max}$. This implies that the Minimax parameters $\{w_i\}$ and $\{\omega_i\}$ for any $R_c > R_c^{\max}$ coincide, since the associated error extremum ϵ is the same in both cases. Figure 5.2 (right) shows the variation of ϵ as a function of $R_c \leq R_c^{\max}$ for different N_q. This plot provides a rough estimation of the error in the final RPA energy for a given pair (R_c, N_q).

More details about the implementation of the minimax quadrature scheme in CP2K are reported in Appendix A.

As shown in Figure 5.1, both approaches, CC and MM, yield exponential convergence, but with a different rate. As a result, for micro-Hartree accuracy, $20 - 30$ quadrature points are required for the CC scheme while $10 - 15$ are sufficient for MM.

Two observations can be made that further improve efficiency. First, $\mathbf{B}^T\mathbf{G}(\omega)\mathbf{B}$ can be favorably rewritten as $\mathbf{B}'(\omega)^T\mathbf{B}'(\omega)$ where $\mathbf{B}'(\omega) = \mathbf{G}(\omega)^{1/2}\mathbf{B}$. This allows us to exploit

symmetry and to replace, in ScaLAPACK notation, PDGEMM by PDSYRK, not only saving half the flops, but equally important, half the memory. Second, $\mathrm{Tr}(\log(\mathbf{1} + \mathbf{Q}(\omega))$ is computed efficiently, in particular avoiding the explicit computation of the logarithm (diagonalization), by exploiting the identity $\mathrm{Tr}[\ln \mathbf{A}] = \ln(\mathrm{Det}[\mathbf{A}])$, where the determinant is obtained from the Cholesky decomposition of the positive definite argument. Such a simple observation leads to a roughly $10\times$ speedup of this $O(N^3)$ step.

Finally, the parallel algorithm is again based on a dual-level parallelization. The first level corresponds to the distribution of the quadrature points over subgroups of processes (the integration groups), while the second level corresponds to the parallel work implied by the matrix multiplication needed for each quadrature point, which is discussed in Section 5.3.3. The redistribution and replication of the $\mathbf{B}$ matrix can be performed efficiently, as the data layout is the same between the integration groups. Since each integration point is essentially independent, parallelization is highly effective also over the quadrature points. Memory requirements restrict the minimum size of an integration group, for example 512 H_2O require at least 4096 nodes per point, so that with 16 integration points this $O(N^4)$ step would likely scale beyond 65536 nodes.

5.3.3 GPU accelerated dense and sparse linear algebra

The $O(N^3)$ and $O(N^4)$ steps of the RI-GPW RPA algorithm are a mixture of sparse and dense linear algebra that benefit significantly from GPU acceleration. During the last year, significant progress has been made in this area, which is summarized here, resulting in all $O(N^4)$ steps being executed efficiently on the GPU.

Sparse Linear algebra: the DBCSR library

As a central part of CP2K, a massively parallel sparse linear algebra library, named DBCSR, has been developed. Its original purpose was linear scaling GGA DFT, enabling calculations with millions of atoms in the condensed phase[98], but it is also useful in the current context of large scale RPA calculations. This multi-year development was needed to accommodate the requirements of the application, and to exploit the internal structure of the matrices that are typical in quantum chemistry. In particular, it provides sparse matrix-matrix multiplication as a key operation, for which it exploits the fact that the sparsity pattern is blocked, with small 'atomic' block sizes that reflect the choice of basis used in the application. The design principles, in particular of its communication scheme, have been summarized in Ref. [99].

Very recently, its GPU implementation has been redesigned and optimized [180]. This implementation provides fully asynchronous operation using double buffered message passing across nodes, and between device and host. Sparse matrix indexing is handled by the host, while most flops are computed on the device, using the CPU as a fall-back. As part of this port, a CUDA library of optimized small matrix matrix multiplication kernels has been generated. This part is essential to deal with the small 'atomic' matrix sub-blocks that are typical for the application, more than 2200 different combinations of block sizes are currently available. The current simulations are characterized by blocks with dimensions that are combinations of 14, 16, and 29. Most important for these sizes is optimal memory transfer within the device, between global, shared and register memory.

An off-line auto-tuning framework determines optimal parameters of several kernel variants, for example, for the 29x29x29 multiplication, 36494 parameter combinations are explored. Within this heuristically pruned search space the median performance is only 160GF. The performance of the optimal kernel found is close to the bandwidth limited peak performance with 387 GF.

Dense linear algebra: libsci_acc

The most important dense linear algebra step is the computation of $\mathbf{B}'(\omega)^T\mathbf{B}'(\omega)$, which is now performed as a parallel multiplication by the ScaLAPACK routine PDSYRK. The latter routine is implemented in Cray's accelerated version of ScaLAPACK, libsci_acc, and ultimately resolves to GPU accelerated DGEMM. However, to obtain good performance, careful tuning, both from within the CP2K code and libsci_acc, was necessary. This step is challenging due to the shape and size of the matrices. Indeed, for the 512 H_2O calculation, the local part of $\mathbf{B}'$ per MPI rank is close to 8Gb. To accommodate this data and libsci_acc's temporary buffers needed for MPI and GPU communication, careful memory management on CP2K's side was needed. Essential for performance is sizing internal ScaLAPACK buffers, determined by a blocking factor (PILAENV), such that all dimensions of the matrices passed to DGEMM are sufficiently large to execute efficiently on the GPU. Host-device-host transfer bandwidth was increased employing page-locked (pinned) memory for buffers and user data. Due to the very 'skinny' shape of the parent matrix (56Mx70K) the resulting data sets were too large to fit in GPU memory, necessitating the adoption of an out-of-core GPU-accelerated DGEMM. Additionally, legacy code in ScaLAPACK, which relied on 32-bit integers for size calculations, had to be fixed. The same skinny nature requires a similar 2D process layout to reach good performance. To illustrate the benefit and progress made from a 'historical' point of view, a 4.4x speedup per integration point is observed in comparing current performance on Daint (see Section 5.5.2) vs. one year old data obtained on Titan (see Section 5.5.1), for 256 H_2O.

5.4 Application: Surprising Water

Water is an abundant liquid covering nearly 70% of the surface of Earth, and constituting a roughly equal fraction of the human body. Yet the properties of water, in its liquid and fourteen other known phases, remain a hotly debated topic [188–192]. Indeed, even the structure of the neat liquid remains disputed, for example if molecules are four-fold coordinated, or if ordinary liquid water is actually a mixture of different liquid phases [193–199]. The pure material must be record-holding with more than 70 anomalous properties being listed, the most famous one being that ice floats on water [200]. However, going beyond the pure form of water is essential. Indeed, water is an important solvent in chemistry. As one example, the folding, unfolding and aggregation of proteins is driven by the water solvent, and the natural equilibrium is easily disrupted by changing the properties of the solvent. The role of water is even more prominent when it is active as an educt in chemical reactions, as shown by two examples. First, the process that removes the anthropogenic CO_2 from the atmosphere is dissolution in the oceans where it reacts with water to form carbonic acid $CO_{2(g)}+H_2O \rightarrow HCO^-_{3(aq)}+H^+_{(aq)}$. This leads to

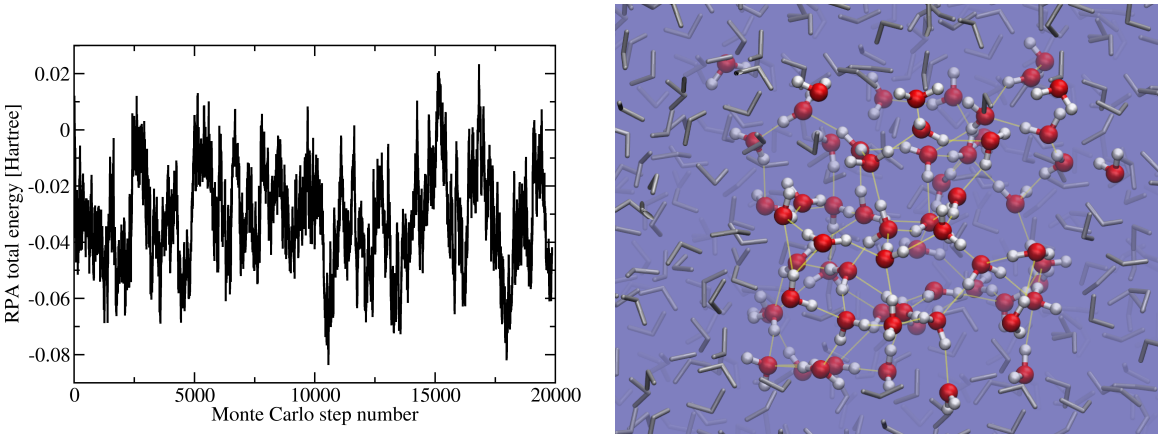


Figure 5.3: Left: For the first time, Monte Carlo simulations of bulk liquid water (64 H_2O) on an RPA potential energy surface can now be performed. Shown is the potential energy evolution, which is characteristic of a well equilibrated system, allowing for a detailed structural analysis. Right: Snapshot of the equilibrated sample of water, with the simulation cell shown embedded in the periodic replicas. Hydrogen bonds, essential for the structure of the liquid and well described by this level of theory, are shown with yellow lines.

acidification, but ultimately also to its removal by mineralization. Second, to power a future hydrogen economy, the carbon neutral approach is electro-chemical water splitting, potentially making water the 'fuel' of the future.

Modeling water has therefore a long tradition. Indeed, already in 1971, soon after the invention of molecular dynamics, Rahman and Stillinger applied this method to liquid water, using a rigid, point charge based description of the molecule [201]. To describe the role of water in chemistry, however, the electronic structure needs to be taken into account. Soon after the development of DFT at its second rung, a pioneering application of the Car-Parrinello technique sampled with a few picoseconds of *ab initio* MD the potential energy surface of 32 water molecules using months of computing power [202]. This was a pioneering application that started a new field. The delicate nature of the system, related to the directional and subtle hydrogen bonding interaction between molecules, can however not be underestimated. Indeed, at the second rung of DFT, the structure of the liquid depends strongly on the choice of functional [203], a feature which is not clearly corrected by 4th rung functionals [104]. 4th rung functionals are needed to obtain a reasonable dielectric constant of ice[204], or to describe the electro-chemistry relevant in water splitting [205]. Even at the 4th rung, van der Waals interactions, crucial for simulations that mimic the common experimental setup where the pressure is constant, are absent, and must be added with empirical force-field like terms [176].

Density functional theory at the 5th rung models includes all these effects in a non-empirical way, and is thus an essential step forward. Up to now, the computational cost of this level excluded extensive simulations. With our recent development of the RI-GPW approach[131], we were for the first time able to demonstrate the excellent performance of an approach including the virtual orbitals, more precisely MP2, for the structure and density of bulk liquid water [177]. As needed for a liquid, finite temperature sampling was employed, in our case by mean of Monte Carlo simulation that require thousands of steps, which makes clear why time-to-solution is a major target for our developments.

The advantage of RPA over MP2 is at least two-fold, first it is non-perturbative and thus applicable to systems with a small gap, and secondly, it displays a favorable $O(N^4)$ instead of $O(N^5)$ scaling, making it far more suitable for large scale applications. Figure 5.3 demonstrates that extended Monte Carlo simulations are indeed possible, yielding a stable and well equilibrated sample. Preliminary analysis shows that both structure and density are in excellent agreement with experiment, but a full analysis will be presented elsewhere [206]. These favorable results, and the computational efficiency with which larger system can be treated, will lead to a routine application of this technique to a broad range of systems.

5.5 System and Environment

All calculations presented have been performed with the CP2K simulation package, on two different hybrid computer architectures, namely Cray XK7 (Titan) and Cray XC30 (Piz Daint). In this way, the impact of important differences between these two architectures can be quantified with a widely used simulation package. At the time of writing, Titan and Piz Daint are the largest supercomputers, as measure by their linpack performance, in the USA and Europe, respectively. However, for the current application, network, memory and compute power are all essential resources. As will be detailed below, Piz Daint is 'minimal in size' to perform the largest calculation presented in this manuscript.

5.5.1 Cray XK7: Titan

Part of the calculations presented in this manuscript have been performed on a Cray XK7 supercomputer, located at the Oak Ridge Leadership Computing Facility (OCLF), named Titan. Titan currently contains 18688 compute nodes linked together with a high-speed interconnect. Each compute node contains one 16-core 2.2GHz AMD Opteron 6274 (Interlagos) processor and 32 GB of RAM, and an NVIDIA Kepler K20X accelerator with 6GB of RAM. Two nodes share a Gemini router, which connects the nodes in a three-dimensional torus. The theoretical peak performance of the system is approximately 27.1 Petaflop/s, with a top500 linpack performance of 17.6 Petaflop/s. Its energy efficiency is 2.1Gflops/W as reported in the green500 list. The total host memory available is 598TB.

5.5.2 Cray XC30: Piz Daint

Part of the calculations presented in this manuscript have been performed on a Cray XC30 supercomputer, located at Swiss National Supercomputing Centre (CSCS), named Piz Daint. Daint currently contains 5272 compute nodes linked together with a high-speed interconnect. Each compute node contains one 8-core 2.6GHz Intel Xeon E5-2670 (Sandy Bridge) processor and 32 GB of RAM, and an NVIDIA Kepler K20X accelerator with 6GB of RAM. Four nodes share an Aries router, which connects the nodes in a dragonfly topology. The rank-3 optical network is fully provisioned, the peak bisection bandwidth is 33TB/s. The theoretical peak performance of the system is approximately 7.8 Petaflop/s, with a top500 linpack performance of 6.3 Petalfop/s. It is also the most energy efficient

system in the linpack Top-10 (3.2Gflops/W, as reported in the green500 list). The total host memory available is 169TB.

5.5.3 Simulation package: CP2K

The calculations in this manuscript are based on the RI-GPW RPA method as implemented in the CP2K simulation package. CP2K is a freely available software package for atomistic simulation [90]. It has a wide range of capabilities, but is best known for its density functional implementation[91], which is often combined with molecular dynamics and Monte Carlo. A comprehensive overview of the major functionality can be found in a recent review, Ref. [207]. CP2K has a large user base and continues to generate a rich scientific output. The web page http://www.cp2k.org/science provides a selection of high profile publications that feature calculations performed with CP2K. Being fully open source, e.g. installable as a package in common Linux distributions and available at computer centers worldwide, precise tracking of the global use is difficult, An indication is given by the fact that the CP2K web page sees approximately 20000 unique visitors per month. On the HPC level, CP2K is the most used simulation code at CSCS, and, for example, the second most used code at the national center in the UK.

The functionality implemented in CP2K does not come without complexity at the source level. Indeed, the code has been actively developed by about a dozen scientists over more than ten years, with an estimated yearly growth in code size of about 80K lines. Porting an application of about 1M lines to a new architecture must be performed piece-wise, and not all code relevant for the current calculations has been ported yet. This 'legacy' part doesn't necessarily imply poor performance. For example, the Hartree-Fock exchange, a key ingredient of 4th rung functionals, was implemented more than five years ago [105] for CPU only, but remains highly competitive. For 1500 atoms (512 H_2O), using 32K cores of Daint, evaluation of the exchange energy takes 12.5s. This still compares favorably to a very recent implementation by a team from IBM, which required 33s per MD step for a slightly smaller system with more than 6M threads on an IBM BG/Q [208].

Finally, CP2K (rev13733) has been built, with local patches that enable memory pinning, using the GNU tool-chain and linked against a customized version of libsci_acc. Both modified versions will be in publically released in the near future. All runs have been performed with 2 MPI ranks per node, which implies four and eight OMP threads per rank on Daint and Titan respectively.

5.6 Performance results

5.6.1 Benchmark system

In order to quantify the performance, timings obtained for RI-GPW RPA energy calculations will be reported for system sizes ranging from 64 H_2O to 512 H_2O. A good quality cc-TZVP basis[131] has been employed for all calculations, and all configurations employ periodic boundary conditions with unit cells that correspond to the experimental density. 32 integration points have been used for the numerical quadrature, yielding converged results. Already the smallest system is comparable to the largest systems reported in literature, while the largest one is 4096 times larger, as estimated from the

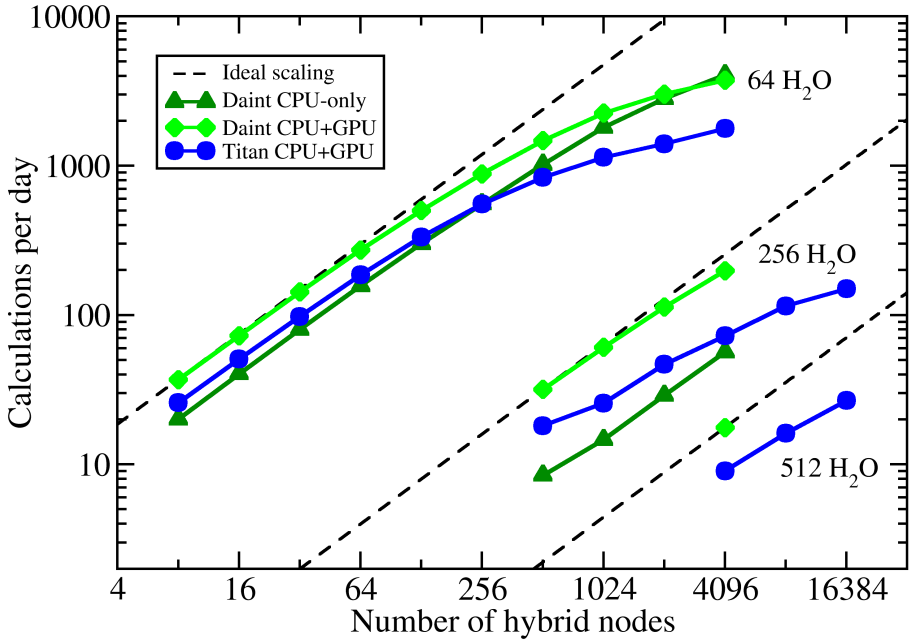


Figure 5.4: Shown is key data from the simulations on Daint (green lines) and Titan (blue lines). 'Calculations per day' (cpd) reflect the ability to do science. At roughly 1000 cpd Monte Carlo simulations become easily possible, at 100 cpd structural relaxation is convenient, while 10 cpd allows for the verification of single point energy differences. These three important targets are easily reached for the three systems presented (64, 256, and 512 H_2O), which span a broad range of computational demands (4096x based on $O(N^4)$ scaling).

$O(N^4)$ scaling of the employed RPA algorithm. In Table 5.1, the sizes of the basis sets and **B** matrix are reported, as well as a conservative estimate of the number of floating point operations needed.

5.6.2 Time to solution

For a methodology to be practical in the field of atomistic simulation, calculations must be completed with weeks, or a few months in the worst case, to meet the rapid pace of scientific progress. In the context of the *ab initio* field, sampling typically requires on the order 20000 energy evaluations to reach 10ps of MD or a similar quality MC simulation. To complete such a simulations within three weeks, an energy evaluation must take less than 1.5 minutes per step, so that roughly 1000 calculation per day (cpd) can be performed. With the presented RI-GPW RPA approach, as shown in Figure 5.4, the target of 1000 cpd is already exceeded on 512 hybrid nodes of Daint for a system of 64 water molecules. This system was previously considered state-of-the-art for single energy evaluations, with 3648 and 8704 primary and auxiliary basis functions respectively. As a result, a new level of simulation methodology has become easily accessible to scientists world-wide. Pushing the limits for this 'small' system, the fastest calculation, obtained on

molecules	64	512
atoms	192	1536
occupied orbitals	256	2048
correlated electrons	512	4096
virtual orbitals	3392	27136
primary basis	3648	29184
auxiliary basis	8704	69632
B dim.	0.9M x 8.7K	55.6M X 69.6K
flop	2.2P	8848.6P

Table 5.1: System sizes for smallest and largest system studied. The flop estimate is an underestimate as only the $O(N^4)$ terms, with their symmetries, are accounted form.

4096 nodes, requires only 21.3 seconds, leading to an impressive 4000+ cpd. To provide context from state-of-the-art calculations in literature, Eshuis and co-workers reported 18.9h for a gas phase RPA calculation with 1117 basis functions, while for condensed phase local MP2 calculations Maschio and co-workers reported timings of 24h for 2884 basis functions [85]. Yet, since the GPU accelerated $O(N^4)$ terms account for only 9.4 and 2.3 seconds on 512 and 4096 nodes, this system is too small to observe the full benefit of the large scale hybrid machines at our disposal. These large hybrid machines thus allow for increasing model size, essentially matching the models commonly used at lower rungs of DFT, a few hundred up to 1500 atoms. Such model sizes become essential to go beyond the pure substance water, for example to study interfaces, but given the $O(N^4)$ scaling of the theory represent an up to 4096x increase in computational demands. For the largest system sizes, we must therefore aim at single point energy calculations, e.g. to verify calculations performed at a lower level, while for intermediate sizes structural relaxation is the target. Defining 10 and 100 calculations per day as the threshold at which this kind of calculations is enabled, we put the bar high compared to the expectations in the field. Using the full size of either Titan or Daint, both targets can be reached for system sizes of 512 and 256 water molecules respectively. The largest system, which is with 1500 atoms in terms of flops roughly 1000x larger than literature calculations and far beyond the reported $O(100)$ electrons limit[179], can be computed in less than 1 hour.

5.6.3 Strong scaling

In this section on strong scaling, the performance of the two different architectures, Cray XC30 and Cray XK7, can be most easily compared. Indeed, this type of simulations, with strong coupling between all tasks, challenges the network. However, first we emphasize that time to solution and strong scaling, as measured by parallel efficiency, go hand in hand, but are not to the same. Indeed, running 64 waters on 64 nodes gives more than 90% parallel efficiency, but with 316s per step would require nearly three months for a trajectory of 20000 steps. By just doubling the computational investment (48% efficiency), the same trajectory could be obtained at 38s per step on 1024 nodes in significantly less than two weeks, an eight-fold improvement in time to solution.

As shown in Figure 5.4, Titan and Daint have rather different performance, which we attribute to the difference in network performance. Two different aspects are relevant.

First, scaling out to larger node counts, the $O(N^4)$ becomes less important, and poorly scaling $O(N^3)$ aspects show up. Prominent is the diagonalization of the Kohn-Sham matrix, which at 16384 nodes is actually more expensive than the $O(N^4)$ terms combined. It does indicate that DFT at the 5th rung, has roughly reached the same time to solution as DFT at the 2nd rung, since they share this common bottleneck. Fortunately, this is an active area of research, with drop-in library replacements being available[209] and in active development[210]. The second aspect is the offset between the scaling curves at low node count. This is directly related to the performance of the PDSYRK call, it is 1.4, 1.7, and 1.8 times faster on Daint than on Titan, for 64, 256, and 512 H_2O respectively. Analysis shows that the underlying DGEMM calls require similar time, as expected since they run on the same GPU, but the MPI calls do not, with a two-fold performance advantage for the Aries over the Gemini network. The importance of data movement is a direct effect of the skinny shape of the matrix, which becomes more pronounced as system size increases. To illustrate this, the $\mathbf{B}'(\omega)^T\mathbf{B}'(\omega)$ multiplication of a 56M x 70K matrix requires the same amount of flops as that of a square 650K x 650K matrix, but requires roughly ten times more data. Hence, linpack performance is only a poor proxy for application performance.

The observed parallel efficiency is nevertheless satisfactory, for 256 H_2O, the parallel efficiency going from 512 to 4096 nodes on Daint is 78% for the overall calculation. The two $O(N^4)$ parts of the code perform at 95% (PDSYRK) and 94% (DBCSR) efficiency. The reason for this is that the parallelization over integration points is highly effective, only beyond 16384 nodes would scaling within the integration group become important. Similar strong scaling is supported by the data obtained on Titan, where 92% efficiency is observed for the $O(N^4)$ terms going from 1024 to 16384 nodes in the case of 256 H_2O, and 95% efficiency is observed in the case of 512 H_2O.

5.6.4 Weak scaling

Weak scaling tests must be performed with a constant ratio of resources to system size. For the current method, these must thus take into account the $O(N^3)$ scaling of the required memory and the $O(N^4)$ scaling of the required flops. Results are favorable. For example, 8, 512 and 4096 nodes are required to fit the calculations in memory for the respective system sizes, perfectly in line with the expected $O(N^3)$ behavior, and confirming 512 molecules is close to the limit also for Titan, since 1024 molecules would need 32768 nodes. For 512 H_2O on 4096 nodes, the user processes require per node 28 GB resident host memory and 5.5 GB device memory, emphasizing the importance of this resource for these calculations. Weak scaling for the runtime requires some extrapolation, as even the runs on Titan (8 to 16384 nodes) do not fully cover the required 4096x range. Runtimes are 6675s (extrapolated), 3359s and 3224s, for 64, 256 and 512 H_2O on 4, 1024 and 16384 nodes, respectively. This represents perfect, or even super-linear weak scaling.

5.6.5 GPU efficiency

Finally, the contribution of GPUs is quantified with runs on Daint with a version of the code and libraries where GPU acceleration is disabled. As only the $O(N^4)$ terms have been ported, the benefit of GPU acceleration on the application level is a moderate 1.8x

for the 64 H_2O system, decreasing as the number of nodes is increased. At 256 H_2O a 3.8x speedup is measured on 512 nodes, retaining 3.5x speedup on 4096 nodes. Focussing only on the $O(N^4)$ terms, sparse and dense linear algebra show 3.5x and 5.5x speedups respectively. Based on the latter two numbers, a conservative estimate for the application level speedup for the 512 H_2O system is 4.4x on 4096 nodes.

5.7 Future systems and science

The calculations performed show the capabilities of current computer systems and hint at what the future will bring. Indeed, models up to 1500 atoms can currently be computed within 1h, or a thousand electrons treated within minutes. In the future, also for such models, time to solution will ultimately be pushed down to less than 1 min. To refer to Figure 5.4, the curve for 256 or 512 H_2O will be extended like we did for 64 H_2O. To reach this target, a 20- or 80-fold speedup over the run on 4096 nodes of Daint will be needed, respectively, which puts this type of calculations firmly on the track to exascale. The data indicates that reaching such goals is rather realistic. Indeed, based on the scaling plots and on algorithmic considerations, already the current combination of hard- and software is likely to show significant speedup up to 65536 nodes, and possibly beyond. Furthermore, the current combination of hard- and software will eventually be upgraded, using the insight gained from this close collaboration as feedback to all parties involved. The requirements of the application as a whole clearly call for a next generation hardware that must be balanced in memory, network and compute. Investments in the full software stack, both system libraries and user code, are required in advance of hardware deployments, to take full advantage of the advanced capabilities.

We have demonstrated the feasibility for RPA calculations of condensed phase systems for systems ranging from 64 to 512 H_2O. The smallest of these systems exceeds what is state-of-the-art for RPA calculations. Yet, most importantly, such calculations can now be performed in seconds rather than hours or days. This enables a new type of scientific calculations, where sampling of condensed phase systems at a finite temperature and using 5th rung functionals becomes easily possible. The RI-GPW RPA approach, implemented in the freely available simulation package CP2K, thus provides the community with a new atomistic simulation tool of improved accuracy and predictivity. Also the larger system sizes can be studied conveniently, so that the prospect of applying this level of theory to more complicated systems, including defects in solids, heterogeneous catalysis, interfaces, etc., is great. We can thus confidently conclude that the importance of the 5th rung of DFT will steadily increase, also in condensed phase simulations.

Chapter 6

Force and Stress in Second Order Møller-Plesset Perturbation Theory for Condensed Phase Systems within the RI-GPW Approach [211]

The forces acting on the atoms as well as the stress tensor are crucial ingredients for calculating the structural properties of systems in the condensed phase. These quantities are obtained as derivatives of the total energy with respect to the atomic coordinates for the former and cell parameters for the latter. In this chapter, the equations for evaluating the derivatives of the second-order Møller-Plesset perturbation energy (MP2) in the framework of the Resolution of Identity Gaussian and Plane Waves (RI-GPW) method are derived. Based on this a massively parallel algorithm has been developed for finite and extended system. The derivatives are evaluated in a way that is fully consistent with how the correlation energy is computed, allowing for energy conserving dynamics in various ensembles. This consistency is non-trivial, given the different ways employed to compute Coulomb, exchange and cannonical four center integrals. The central aspect of the RI-GPW approach is the dual representation of the RI fitting densities in term of Gaussian and Plane Waves auxiliary functions. This allows for the treatment of the Coulomb interactions that is particularly efficient in the condensed phase, involving the evaluation of integrals between Gaussian basis functions and the electrostatic potential associated to each RI density. The designed parallel algorithm displays, with respect of the system size, cubic, quartic and quintic requirements respectively for the memory, communication and computation. All these requirements are reduced with an increasing number of processes and the measured performance shows excellent parallel scalability and efficiency up to thousands of nodes. In addition to that, the computationally more demanding part, that is the quintic scaling steps, can be accelerated by employing graphics processing units (GPU) showing, for large systems, a gain of almost a factor two compared to the standard only CPU case. In this way the evaluation of the derivatives of the RI-MP2 energy can be performed within minutes employing a few hundred hybrid nodes for systems containing hundreds of atoms and thousands of basis functions. Geometry optimization as well as full cell relaxation have been performed for a variety of molecular crystals including NH_3, CO_2, Formic Acid and Benzene.

6.1 Introduction

The energy evaluated with the second-order Møller-Plesset (MP2) perturbation theory represents an effective way to improve the Hartree-Fock (HF) ground state by including electron correlation effects [13, 14]. In this respect, MP2, also referred as second-order many body perturbation theory (MBPT(2)), offers many appealing features, such as size consistency and the capability to correctly account for dispersion interactions [30]. In addition to that, MP2 is an *ab-initio* method that can accurately describe hydrogen-bond, covalent and ionic interactions from first principles. Moreover, among the electronic structure methods, MP2 is probably that one displaying the simplest and more compact form. For these reasons MP2 is often used as a reference for testing and benchmarking new approximate methods and MP2-like correlation has also been included in Density Functional Theory (DFT) with the introduction of Double-Hybrid Density Functionals [17, 19]. However, despite the advantages of MP2, there are two flaws that have posed limitations to its spread use. The first is related to the unfavorable $O(N^5)$ growth of the computational effort with respect to the systems size. The second, common for all wavefunction based correlation method, is the slow convergence of the MP2 energy with respect to the basis set size, due to the electron coalescence cusp [120, 134]. In order to extend the applicability of MP2 to large system these limitations have to be overcome [52].

Many approaches have been proposed in order to reduce the formal $O(N^5)$ scaling and they can be classified as Laplace-Transformed MP2, [48, 62–67, 132] Local MP2 (LMP2), [49, 53–61] and Stochastic [212–215] methods, while explicitly correlated schemes can be used for accelerating the convergence of the MP2 energy with respect to basis set size (F12-MP2) [74, 216, 217]. Furthermore, the Resolution of Identity (RI) [46, 47, 68, 71–73, 127, 137, 141] approximation, sometimes referred as Density Fitting (DF), has shown to greatly speed up the evaluation of the MP2 energy giving almost a order of magnitude reduction of the computational cost without significant loss of accuracy [69, 70, 218]. It is also important to note that parallel computing has become of prime importance in quantum chemistry as a tool for reducing the time to solution for these calculations. In this respect many parallel algorithms have been proposed [44, 75–85, 131, 139] showing an efficiency growing at the same footstep as the increase of the computational power.

Thanks to all these improvements the applicability of the MP2 theory has been stretched continuously over the time and recently an RI-MP2 Monte Carlo (MC) simulation under ambient conditions of bulk liquid water has been reported [177]. The advantage of the MC scheme is that only the total energy is required in order to calculate the ensemble averages. On the other hand, an efficient MC sampling needs a sufficient knowledge of the system under study necessary to define "smart" trial move. This makes the use of MC method less straightforward than *e.g* Molecular Dynamics (MD), for which the ensemble averages are obtained by integrating the classical equations of motion. In this case the forces acting on atoms have to be computed, obtained from the derivative of the total energy with respect to the atomic positions. Furthermore, MC does not give access to truely dynamical properties, i.e. derived from time correlation functions, such as for example, diffusion constants and vibrational spectroscopy. To obtain those, accurate energy conserving (NVE) simulations have to be performed, requiring consistent forces.

The evaluation of the derivatives at the MP2 level is more intricate compared to their

computation at the HF level. This is because, contrary to the HF case, the correlation energy obtained from perturbation theory is non stationary with respect to the molecular orbital expansion coefficients, implying that first order orbital response has to be computed. The theory and equations for calculating the energy derivatives at the MP2 [219–222] and RI-MP2 [147] level have been derived and reported by many authors, together with many serial [223–225] and parallel [226–229] implementations.

Here the equations for evaluating the derivatives of the second-order Møller-Plesset perturbation energy in the framework of the Resolution of Identity Gaussian and Plane Waves (RI-GPW) are presented. The derivatives are evaluated consistently to the way the RI-MP2 energy is computed, [131] and are of general validity for both finite and extended systems. The central idea in the RI approximation is the introductions of an atom-center Gaussian auxiliary basis used for fitting pairwise products of atomic orbital basis functions. In addition to the representation in term of Gaussian functions, the RI-fitting densities within the RI-GPW method are expressed also employing an auxiliary basis of Plane Waves (PW). This choice allows for rapid conversion between direct and reciprocal space representation of the density by employing fast Fourier transformations (FFT). In this way the treatment of the Coulomb interactions is efficiently accomplished by integration of the electrostatic potential associated to each RI-fitting density over the pairs of primary basis functions, where the electrostatic potential is obtained in a plane wave basis set after the solution of the Poisson equation in Fourier space. As a drawback, the GPW method requires smooth densities, implying that pseudopotentials have to be employed. All-electron calculations are possible within the Gaussian and Augmented Plane Wave (GAPW) scheme [96, 97], however the actual implementation is currently limited to the GPW method only.

An implementation of the analytical energy gradients at the MP2 level for extended systems has been reported by Hirata and coworkers [89]. In this case the formulation is based on the crystal orbital theory implying that two-electron integrals are obtained by k-point sampling in the first Brillouin zone. Moreover the applications are limited to polymers (periodic 1D) with small basis. The difference compared to the method presented here relies in the way the two-electron integrals are computed. In fact in GPW, the sampling of the first Brillouin zone is restricted to the Gamma point only. The GPW approach converges to the same value as obtained from full k-point sampling if a sufficiently large supercell is chosen.

For the presented scheme a massively parallel algorithm has been designed and implemented in CP2K[90]. The parallel algorithm displays, with respect of the system size, cubic, quartic and quintic effort respectively for the memory, communication and computation. All these requirements scale increasing the number of processes and the measured performance displays excellent parallel scalability and efficiency up to thousands of nodes. Moreover, in the actual implementation the computationally more demanding part, that is the quintic scaling steps, can be accelerated by employing graphics processing units (GPU). Compared to the standard only CPU case, this leads, in general, to a speed-up of a factor grater than 4 for the $O(N^5)$ parts of the algorithm, resulting, for the largest cases, in an almost factor 2 reduction in the overall time for the calculation.

Several benchmark calculations have been reported with a particular focus on molecular crystals including NH_3, CO_2, Formic Acid and Benzene. In general it has been observed that the effort for the calculation of the derivatives at the RI-MP2 level is between 4 to 5

times more expensive than computing only the energy.

6.2 Theory

In this section the basic equations necessary for implementing the first derivatives of the RI-MP2 energy are briefly presented referring to the original works for more details [147, 222, 225, 229]. More information are reported in the cases for which the general theory is combined with the GPW approach. The following index notation has been adopted: $i, j, k, \ldots$ refer to canonical occupied molecular orbitals (MOs), $a, b, c, \ldots$ to canonical virtual MOs, $p, q, r, \ldots$ to general canonical MOs, $\mu, \nu, \lambda, \ldots$ to primary atomic orbital basis set functions (AO), $P, Q, R, \ldots$ to auxiliary AO basis set functions (AUX). The one electron MO, primary AO and auxiliary AO functions are symbolized respectively with ψ, ϕ and χ. The number of occupied and virtual orbitals is denoted by o and v, while the total number of primary and auxiliary basis functions as n and N_a. In order to express, in general, the system size, the symbol N is used. Given a perturbation parameter x, *e.g.* a nuclear displacement, the superscript ^{x} represents the derivative with respect to x, while $^{(x)}$ denotes the skeleton derivative, that is derivatives of the AO integrals only (*i.e.* without considering the derivatives of the expansion coefficients of the MOs).

The MP2 Energy within the RI-GPW Method

In Second Order Møller-Plesset perturbation theory, the correlation energy $E^{(2)}$ for a closed shell restricted Hartree-Fock reference wave function is given by:

$$E^{(2)} = -\sum_{i \leq j}^{o} (2 - \delta_{ij}) \sum_{ab}^{v} \frac{(ia|jb)[2(ia|jb) - (ib|ja)]}{\epsilon_a + \epsilon_b - \epsilon_i - \epsilon_j} \tag{6.1}$$

where ϵ_p are orbital energies, δ_{ij} is the Kronecker delta and $(ia|jb)$ is a two-electron repulsion integral (ERI) over MO in Mulliken notation

$$(ia|jb) = \int\int \psi_i(\vec{r}_1)\psi_a(\vec{r}_1)\frac{1}{\vec{r}_{12}}\psi_j(\vec{r}_2)\psi_b(\vec{r}_2)d\vec{r}_1 d\vec{r}_2. \tag{6.2}$$

In a standard canonical MP2 energy algorithm the computation of the $(ia|jb)$ integrals is performed via four consecutive integral transformations of the ERIs over AO $(\mu\nu|\lambda\sigma)$:

$$(ia|jb) = \sum_{\mu} C_{\mu i} \sum_{\nu} C_{\nu a} \sum_{\lambda} C_{\lambda j} \sum_{\sigma} C_{\sigma b}(\mu\nu|\lambda\sigma), \tag{6.3}$$

where the $C_{\kappa p}$ represent elements of the MO coefficient matrix. Each of the four quarter transformations has a formal computational effort that grows as $O(N^5)$ that eventually reflects into the asymptotic scaling associated to the evaluation of MP2 energy [148].

The resolution of identity approximation [144, 145] is an effective technique that allows to accelerate the evaluation of the $(ia|jb)$ ERIs. It consists in the introduction of an auxiliary Gaussian basis set $\{\chi_P\}$ used to factorize the $(ia|jb)$ integrals according to:

$$(ia|jb)_{\mathrm{RI}} = \sum_{P} B_{ia}^{P} B_{jb}^{P}, \tag{6.4}$$

where $\mathbf{B}$ is a matrix with ov rows and N_a columns given by:

$$B_{ia}^{P} = \sum_{Q} (ia|Q) V_{QP}^{-1/2} \tag{6.5}$$

and $V_{QP}^{-1/2}$ are the matrix elements of the inverse square root of the Coulomb metric [146] $V_{QP} = (Q|P)$

$$(Q|P) = \int\int \chi_Q(\vec{r}_1) \frac{1}{|\vec{r}_1 - \vec{r}_2|} \chi_P(\vec{r}_2) d\vec{r}_1 d\vec{r}_2. \tag{6.6}$$

Since the three center integrals $(ia|Q)$ are computed starting from integrals over AO

$$(\mu\nu|Q) = \int\int \phi_\mu(\vec{r}_1)\phi_\nu(\vec{r}_1) \frac{1}{|\vec{r}_1 - \vec{r}_2|} \chi_Q(\vec{r}_2) d\vec{r}_1 d\vec{r}_2, \tag{6.7}$$

the final expression for the B_{ia}^{P} elements reads:

$$B_{ia}^{P} = \sum_{\nu} C_{\nu a} \sum_{\mu} C_{\mu i} \sum_{Q} (\mu\nu|Q) V_{QP}^{-1/2}. \tag{6.8}$$

The RI approximation to the ERIs over MO paves many practical advantages:

- The $(ia|jb)_{\mathrm{RI}}$ ERIs can be evaluated without significant loss of accuracy even employing an auxiliary basis that is only $2-4$ times larger than the primary basis [69, 147, 218].
- The effort for the integral computation is strongly reduced since 4 index integrals over AO $(\mu\nu|\lambda\sigma)$ are never generated and only three $(\mu\nu|Q)$ and two $(Q|P)$ center ERIs are required. This means that, the integral computation requires formally $O(N^3)$ operations while the integral transformations (Equation (6.8)) scale asymptotically as $O(N^4)$.
- As shown by Equation (6.4), the whole set of four index ERIs over MO $(ia|jb)_{\mathrm{RI}}$ can be efficiently evaluated from the three index intermediates B_{ia}^{P} completely by matrix-matrix multiplications.
- Since for the generation of the $(ia|jb)_{\mathrm{RI}}$ only the matrix $\mathbf{B}$ has to be stored, the required memory grows as $O(N^3)$.

The application of the RI approximation to the MP2 energy calculation is straightforward [68]. It consists simply in the replacement of the $(ia|jb)$ integrals in Equation (6.1) with the approximated $(ia|jb)_{\mathrm{RI}}$ given in Equation (6.4). The computation of the $(ia|jb)_{\mathrm{RI}}$ requires $O(o^2v^2N_a)$ operations implying that the RI-MP2 method is also scaling $O(N^5)$. Thus the advantage of RI-MP2, compared to a standard MP2 implementation, relies on the reduced required memory and prefactor associated to the computation of the $(ia|jb)$ via Equation (6.4) instead of Equation (6.3).

According to what shown so far, it appears clear that applying the RI approximation to the MO-ERIs requires the computation of three $(\mu\nu|Q)$ and two $(Q|P)$ center ERIs. In particular for condensed phase systems, for which periodic boundary conditions (PBC) are mandatory, these intermediates have to account for the requirement that the simulation

cell is infinitely replicated in all directions in space. In order to accomplish this task, the Gaussian and Plane-Waves method [86, 91] has been extended for handling integrals that arise in wave-function correlation methods [44, 131].

In general, the GPW method is an efficient approach for treating Coulomb interactions between Gaussian basis elements and any electrostatic density ρ that fulfill the PBC of the considered system:

$$(\mu\nu|\rho) = \int\int \phi_\mu(\vec{r}_1)\phi_\nu(\vec{r}_1)\frac{1}{|\vec{r}_1 - \vec{r}_2|}\rho(\vec{r}_2)d\vec{r}_1 d\vec{r}_2. \tag{6.9}$$

Here, if ρ is considered as the total electronic density, then the form of the last equation is essentially identical to the one used to compute matrix elements of the Hartree potential [86]. Thus, the highly efficient implementation of that operation in CP2K[90] can be directly used and we refer to Ref. [91] for a detailed discussion.

In particular for the RI case, focusing on three center integrals, they are computed, Equation (6.8), starting from the integrals over AOs that are subsequently transformed with the two matrices $\mathbf{C}$ and $\mathbf{V^{-1/2}}$. Employing the GPW method, Equation (6.9), the index transformation over the auxiliary basis can be avoided, since it is possible to directly compute half transformed integrals for an associated density ρ^P as

$$\begin{aligned} B^P_{\mu\nu} &= \sum_Q (\mu\nu|Q)V^{-1/2}_{QP} \\ &= \int\int \phi_\mu(\vec{r}_1)\phi_\nu(\vec{r}_1)\frac{1}{\vec{r}_{12}}\left[\sum_Q \chi_Q(\vec{r}_2)V^{-1/2}_{QP}\right]d\vec{r}_1 d\vec{r}_2 \\ &= \int \phi_\mu(\vec{r}_1)\phi_\nu(\vec{r}_1)\left[\int \frac{\rho^P(\vec{r}_2)}{\vec{r}_{12}}d\vec{r}_2\right]d\vec{r}_1 \\ &= \int \phi_\mu(\vec{r}_1)\phi_\nu(\vec{r}_1)v^P(\vec{r}_1)d\vec{r}_1. \end{aligned} \tag{6.10}$$

The same approach holds for the $(P|Q)$ integrals with the only difference that the potential is calculate from the density associated to a single Gaussian auxiliary basis function.

The key aspect in GPW is that the density ρ^P is expressed on a regular grid, or, in equivalent terms, ρ^P is expanded in an auxiliary basis of plane waves (PW)

$$\rho^P(\vec{R}) \approx \frac{1}{\Omega}\sum_{\vec{G}} \rho^P(\vec{G})e^{i\vec{G}\cdot\vec{r}} \tag{6.11}$$

where $\rho^P(\vec{G})$ are the Fourier coefficients of the density, Ω is the volume of the simulation cell and the sum over the reciprocal lattice vectors $\vec{G}$ is determined by the size S of the PW basis. Fast Fourier transforms (FFTs) allow for switching representation between real space ($\rho^P(\vec{R})$) and reciprocal ($\rho^P(\vec{G})$) space with an associated computational effort that grows only as $O(S\log S)$. In this way, the electrostatic potential v^P in Equation (6.10) can be efficiently obtained in a plane waves basis set after solving the Poisson equation in Fourier space

$$v^P(\vec{G}) = \frac{4\pi}{G^2}\rho^P(\vec{G}), \tag{6.12}$$

while an additional back FFT (FFT^{-1}) will yield the potential in real space.

An extensive description, together with the implementation details, of the RI-GPW method can found in Ref. [131], here only the most important features are highlighted:

- The accuracy of the calculated $B^P_{\mu\nu}$ integrals in Equation (6.10) can be systematically improved by increasing the PW basis set size (resolution of the grid) [44]. This is conventionally done by specifying the energy cutoff that limits the kinetic energy of the PWs.
- PW auxiliary basis is a natural choice for periodic systems, but it can equally be used for gas phase or surface calculations [93–95].
- All-electron calculations are not possible, and that pseudopotentials have to be employed [101].
- For each electrostatic potential v^P all matrix elements that are non-zero within a given threshold (ϵ_{grid}) can be obtained in linear scaling time [91].
- $B^P_{\mu\nu}$ are transform from the AO basis to the MO basis (B^P_{ia}) via two consecutive matrix-matrix multiplications, $\mathbf{B}^P_{\mathrm{MO}} = \mathbf{C}_o^\dagger \mathbf{B}^P_{\mathrm{AO}} \mathbf{C}_v$, with $\mathbf{C}_o$ and $\mathbf{C}_v$ being respectively the occupied and virtual parts of the coefficient matrix. The multiplication by $\mathbf{C}_o$ can exploit the sparsity of $\mathbf{B}^P_{\mathrm{AO}}$, implying an $O(no)$ scaling for each P, while the final multiplication can not exploit sparsity and is asymptotically dominant, scaling as $O(onv)$.

The Analytic Derivatives for RI-MP2

The analytic derivative of the RI-MP2 energy $E^{(2)}_{\mathrm{RI}}$ with respect to a perturbation parameter x, for a closed-shell restricted Hartree-Fock wave function, is given by [147, 225, 229]:

$$E^{(2)^x}_{\mathrm{RI}} = \frac{dE^{(2)}_{\mathrm{RI}}}{dx} = 4 \sum_{Q}^{\mathrm{AUX}} \sum_{\mu\nu}^{\mathrm{AO}} \Gamma^Q_{\mu\nu} (\mu\nu|Q)^x - 2 \sum_{PQ}^{\mathrm{AUX}} \Gamma^{PQ} (P|Q)^x + 2 \sum_{pq}^{\mathrm{MO}} \left[P^{(2)}_{pq} F^{(x)}_{pq} - W^{(2)}_{pq} S^{(x)}_{pq} \right]. \tag{6.13}$$

In the above expression, for each summation, a common structure can be recognized, that is the contraction of terms involving AO derivatives $(\mu\nu|Q)^x, (P|Q)^x, F^{(x)}_{pq}, S^{(x)}_{pq}$, with elements of the intermediates $\Gamma^Q_{\mu\nu}, \Gamma^{PQ}, P^{(2)}_{pq}, W^{(2)}_{pq}$. The contribution to the derivatives of $E^{(2)}_{\mathrm{RI}}$ coming from the first two summations, referred as non-separable part, is specific to the RI-MP2 method. It involves the contraction of 3- and 2-center RI integral derivative $(\mu\nu|Q)^x, (P|Q)^x$ with 3- and 2-index quantities namely non-separable correction to the 2-particle density matrix (2-PDM), $\Gamma^Q_{\mu\nu}$ and Γ^{PQ}. These two specific quantities are given by:

$$\Gamma^Q_{\mu\nu} = \sum_{i}^{occ} C_{i\mu} \sum_{a}^{virt} C_{\nu a} \Gamma^Q_{ia} \tag{6.14}$$

$$\Gamma_{ia}^{Q} = \sum_{P}^{\text{AUX}} Y_{ia}^{P} V_{PQ}^{-1/2} \tag{6.15}$$

$$Y_{ia}^{P} = \sum_{jb} \frac{2(ia|jb) - (ib|ja)}{\epsilon_i + \epsilon_j - \epsilon_a - \epsilon_b} B_{jb}^{P} \tag{6.16}$$

$$\Gamma^{PQ} = \sum_{R}^{\text{AUX}} \sum_{ia} \Gamma_{ia}^{P} B_{ia}^{R} V_{RQ}^{-1/2} \tag{6.17}$$

Once $\Gamma_{\mu\nu}^{Q}$ and Γ^{PQ} are made available, the non-separable contribution to $E_{\text{RI}}^{(2)^x}$ is obtained by contraction with $(\mu\nu|Q)^x$ and $(P|Q)^x$, which are computed consistently to $(\mu\nu|Q)$ and $(P|Q)$ by employing the same GPW scheme. This leads, for the 3-center case, to:

$$\begin{aligned}
\sum_{Q}^{\text{AUX}} \sum_{\mu\nu}^{\text{AO}} \Gamma_{\mu\nu}^{Q} (\mu\nu|Q)^x &= \sum_{Q}^{\text{AUX}} \sum_{\mu\nu}^{\text{AO}} \Gamma_{\mu\nu}^{Q} \left[(\mu^x\nu|Q) + (\mu\nu^x|Q) \right] + \sum_{Q}^{\text{AUX}} \left(\sum_{\mu\nu}^{\text{AO}} \Gamma_{\mu\nu}^{Q} \phi_\mu \phi_\nu | Q^x \right) \\
&= \sum_{Q}^{\text{AUX}} \sum_{\mu\nu}^{\text{AO}} \Gamma_{\mu\nu}^{Q} \left[(\mu^x\nu|\rho^Q) + (\mu\nu^x|\rho^Q) \right] + \sum_{Q}^{\text{AUX}} (\rho^{\Gamma^Q}|Q^x) \\
&= \sum_{Q}^{\text{AUX}} \sum_{\mu\nu}^{\text{AO}} \Gamma_{\mu\nu}^{Q} \int \left[\phi_\mu^x(\vec{r}) \phi_\nu(\vec{r}) + \phi_\mu(\vec{r}) \phi_\nu^x(\vec{r}) \right] v_H^Q(\vec{r}) d\vec{r} \\
&\quad + \sum_{Q}^{\text{AUX}} \int \chi_Q^x(\vec{r}) v_H^{\Gamma^Q}(\vec{r}) d\vec{r}
\end{aligned} \tag{6.18}$$

where $v_H^{\Gamma^Q}$ is the electrostatic potential related to the $\sum_{\mu\nu}^{\text{AO}} \Gamma_{\mu\nu}^{Q} \phi_\mu(\vec{r}) \phi_\nu(\vec{r})$ density, while v_H^Q is the counterpart associated to the single auxiliary Gaussian function $\chi_Q(\vec{r})$. For the 2-center case, exploiting the fact that Γ^{PQ} is symmetric, the analogous approach gives:

$$\sum_{PQ}^{\text{AUX}} \Gamma^{PQ} (P|Q)^x = 2 \sum_{PQ}^{\text{AUX}} \Gamma^{PQ} \int \chi_P^x(\vec{r}) v_H^Q(\vec{r}) d\vec{r} = 2 \sum_{Q}^{\text{AUX}} \int \chi_Q^x(\vec{r}) v_H^{\Gamma^Q}(\vec{r}) d\vec{r} \tag{6.19}$$

where $v_H^{\Gamma^Q}$ is the potential obtained from the $\sum_{P} \Gamma^{PQ} \chi_P(\vec{r})$ density, while v_H^Q is the same as in Equation (6.18). The two formulations given in Equation (6.19) are equivalent and, even if the latter offers more advantages in term of computational efficiency, the former is preferred since v_H^Q can be subsequently reused in Equation (6.18).

The last summation in Equation (6.13) consists in the contraction of $P_{pq}^{(2)}$, the MP2 correction to the 1-particle density matrix (1-PDM), and $W_{pq}^{(2)}$, the MP2 correction to the energy-weighted density matrix, with the skeleton derivatives of the Fock and overlap matrix elements

$$F_{pq}^{(x)} = \sum_{\mu\nu}^{\text{AO}} C_{\mu p} \left[h_{\mu\nu}^x + \sum_{\lambda\sigma} P_{\lambda\sigma}^{\text{HF}} (\mu\nu|\lambda\sigma)^x - \frac{1}{2} \sum_{\lambda\sigma} P_{\lambda\sigma}^{\text{HF}} (\mu\lambda|\nu\sigma)^x \right] C_{\nu q} \tag{6.20}$$

$$S_{pq}^{(x)} = \sum_{\mu\nu}^{\mathrm{AO}} C_{\mu p} S_{\mu\nu}^{x} C_{\nu q}. \tag{6.21}$$

In Equation (6.20), $h_{\mu\nu}^{x}$ and $(\mu\nu|\lambda\sigma)^{x}$ are respectively the derivatives of the one-electron Hamiltonian integrals and the 4-index ERIs over AO, while $P_{\mu\nu}^{\mathrm{HF}} = 2\sum_{i}^{occ} C_{\mu i} C_{\nu i}$ is the Hartree-Fock density matrix obtained from the converged SCF procedure. In order to take advantages from sparsity, the update of the $E_{\mathrm{RI}}^{(2)}$ derivative is performed in the AO basis, previous back transformation of $P_{pq}^{(2)}$ and $W_{pq}^{(2)}$ from the MO basis.

In the framework of the GPW method, $h_{\mu\nu}^{x}$ contains the derivative of the matrix element of the electronic kinetic energy, short range part of the local pseudopotential and the non-local pseudopotential. These terms are computed analytically and explicit formulas can be found in Ref. [91]. The exact HF exchange contributions (last summation inside the squared brake in Equation (6.20)), are calculated consistently, via 4-index ERI derivatives, with the Γ-point implementation based on a short range (truncated) Coulomb operator in the case of PBC [104, 105]. Due to the dual representation of the density in GPW, special care has to be taken for the derivative of the Hartree matrix elements (second term inside the squared brake in Equation (6.20)). In particular, it is convenient to reformulate the contribution coming from the Hartree energy in terms of electrostatic densities, this is accomplished by exploiting the symmetry of $P_{\mu\nu}^{\mathrm{HF}}$, $P_{\mu\nu}^{(2)}$ and $(\mu\nu|\lambda\sigma)$ derivatives:

$$\begin{aligned}
&\sum_{\mu\nu\lambda\sigma} P_{\mu\nu}^{(2)} P_{\lambda\sigma}^{\mathrm{HF}} (\mu\nu|\lambda\sigma)^{x} = \\
&\quad = 2\sum_{\mu\nu} P_{\mu\nu}^{(2)} (\mu^{x}\nu|\sum_{\lambda\sigma} P_{\lambda\sigma}^{\mathrm{HF}} \phi_\lambda\phi_\sigma) + 2\sum_{\lambda\sigma} P_{\lambda\sigma}^{\mathrm{HF}} (\sum_{\mu\nu} P_{\mu\nu}^{(2)} \phi_\mu\phi_\nu|\lambda^{x}\sigma) \\
&\quad = 2\sum_{\mu\nu} P_{\mu\nu}^{(2)} (\mu^{x}\nu|\rho^{\mathrm{HF}}) + 2\sum_{\lambda\sigma} P_{\lambda\sigma}^{\mathrm{HF}} (\rho^{(2)}|\lambda^{x}\sigma) \\
&\quad = 2\sum_{\mu\nu} P_{\mu\nu}^{(2)} \int \phi_\mu^{x}(\vec{r})\phi_\nu(\vec{r}) v_H^{\mathrm{HF}}(\vec{r})d\vec{r} + 2\sum_{\lambda\sigma} P_{\lambda\sigma}^{\mathrm{HF}} \int \phi_\lambda^{x}(\vec{r})\phi_\sigma(\vec{r}) v_H^{(2)}(\vec{r})d\vec{r}
\end{aligned} \tag{6.22}$$

where the v_H^{HF} and $v_H^{(2)}$ are the Hartree potentials associated to $\rho^{\mathrm{HF}} = \sum_{\lambda\sigma} P_{\lambda\sigma}^{\mathrm{HF}} \phi_\lambda(\vec{r})\phi_\sigma(\vec{r})$ and $\rho^{(2)} = \sum_{\mu\nu} P_{\mu\nu}^{(2)} \phi_\mu(\vec{r})\phi_\nu(\vec{r})$. For the sake of completeness, it has to be mentioned that in the GPW scheme the total Hartree potential includes an additional term that comes from the introduction of a Gaussian charge distribution at each nucleus $\rho_c(\vec{r})$. This is commonly done in Ewald sum method in order to decouple the long and short range treatment of the electrostatic interactions. The contribution to $E_{\mathrm{RI}}^{(2)^{x}}$ associated to this additional term is accounted by integrating the $\rho_c(\vec{r})$ derivative with the previously defined $v_H^{(2)}$ potential

$$2\int \rho_c^{x}(\vec{r}) v_H^{(2)}(\vec{r})d\vec{r} \tag{6.23}$$

as likewise done for the similar term in standard HF method [91]. For efficiency reasons, in order to avoid the recomputations of integrals derivatives, the contraction of $P_{\mu\nu}^{(2)}$ and

$W^{(2)}_{\mu\nu}$ is performed, when possible, at the same time with $P^{\mathrm{HF}}_{\mu\nu}$ and $W^{\mathrm{HF}}_{\mu\nu}$ matrices, *i.e.* simultaneously during the calculation of the HF energy derivatives.

At this stage, the only missing quantities that remain to be defined are $P^{(2)}_{pq}$ and $W^{(2)}_{pq}$. These matrices are usually calculated in the MO basis, and they are the result of the composition of terms that have a different definition according to which block of the matrix they refer, namely occupied-occupied (occ-occ), virtual-virtual (virt-virt) and occupied-virtual (occ-virt). Concerning the MP2 correction to the 1-PDM $P^{(2)}_{pq}$, the occ-occ and virt-virt blocks are defined as:

$$P^{(2)}_{ij} = -\sum_{ab}^{virt}\sum_{k}^{occ} t^{ab}_{ik}\frac{(ja|kb)}{\epsilon_j + \epsilon_k - \epsilon_a - \epsilon_b} \tag{6.24}$$

$$P^{(2)}_{ab} = \sum_{ij}^{occ}\sum_{c}^{virt} t^{ac}_{ij}\frac{(ib|jc)}{\epsilon_i + \epsilon_j - \epsilon_b - \epsilon_c} \tag{6.25}$$

where t^{ab}_{ij} are the MP2 amplitudes, that in the restricted closed shell HF case take the form:

$$t^{ab}_{ij} = \frac{2(ia|jb) - (ib|ja)}{\epsilon_i + \epsilon_j - \epsilon_a - \epsilon_b}. \tag{6.26}$$

The virt-occ block of $P^{(2)}$ contains information related to the orbital relaxation caused by the perturbation x, [220] (*i.e.* first order response of the MO coefficients) [230]. It is computed as the solution of the Z-vector equations [231]

$$\sum_{a}^{virt}\sum_{i}^{occ}\left[\delta_{ij}\delta_{ab}(\epsilon_a - \epsilon_i) + A_{aibj}\right] P^{(2)}_{ai} = -L_{bj} \tag{6.27}$$

where A_{aibj} is an element of the orbital Hessian matrix

$$A_{aibj} = 4(ai|bj) - (ab|ij) - (aj|bi), \tag{6.28}$$

and L is a specific RI-MP2 Lagrangian matrix given by:

$$L_{bj} = 2\sum_{a}^{virt}\sum_{Q}^{\mathrm{AUX}}(ba|Q)\Gamma^{Q}_{ja} - 2\sum_{i}^{occ}\sum_{Q}^{\mathrm{AUX}}(ij|Q)\Gamma^{Q}_{ib} + \sum_{ac}^{virt} P^{(2)}_{ac}A_{acbj} + \sum_{ik}^{occ} P^{(2)}_{ik}A_{ikbj}. \tag{6.29}$$

The first two terms in Equation (6.29), namely $L_{bj}(1)$ and $L_{bj}(2)$, are computed within the mixed Lagrangian formalism [225], that is, starting from the counterpart $L_{\mu j}(1)$ and $L_{b\nu}(2)$ in a mixed AO/MO basis

$$L_{jb}(1) = \sum_{\mu}^{\mathrm{AO}} C_{\mu b}L_{\mu j}(1) \tag{6.30}$$

$$L_{jb}(2) = \sum_{\nu}^{\mathrm{AO}} L_{b\nu}(2)C_{\nu j} \tag{6.31}$$

where:

$$L_{\mu j}(1) = 2 \sum_{\nu}^{\text{AO}} \sum_{Q}^{\text{AUX}} (\mu\nu|Q)\Gamma^{Q}_{j\nu} \tag{6.32}$$

$$L_{b\nu}(2) = -2 \sum_{i}^{occ} \sum_{Q}^{\text{AUX}} (i\nu|Q)\Gamma^{Q}_{ib}. \tag{6.33}$$

The above reformulation, allows to accumulate the contributions to $L_{\mu j}(1)$ and $L_{b\nu}(2)$ together with the contraction of $\Gamma^{Q}_{\mu\nu}$ and Γ^{PQ} with the integrals derivatives $(\mu\nu|Q)^x$ and $(P|Q)^x$. This choice is thus particularly convenient from a computational standpoint, since many intermediates, such as $\Gamma^{Q}_{j\nu}$, are in common for both updates and don't need to be recomputed. Moreover, the $(\mu\nu|Q)$ integrals have to be recalculated and this is performed at the same time with the computation of the corresponding derivatives, allowing a further save since all grid operations, such as FFT's, are performed once for both terms.

The calculation of the off-diagonal elements of $P^{(2)}_{ij}$ and $P^{(2)}_{ab}$, defined in Equation (6.24) and (6.25), can be equivalently computed (within a canonical reformulation [147, 229, 232]) from $L_{ij}(1) = \sum_{\mu} C_{\mu j} L_{\mu i}(1)$ and $L_{ab}(2) = \sum_{\nu} L_{a\nu}(2) C_{\nu b}$ as:

$$P^{(2)}_{ij} = \frac{1}{2} \frac{L_{ij}(1) - L_{ji}(1)}{\epsilon_j - \epsilon_i} \tag{6.34}$$

$$P^{(2)}_{ab} = -\frac{1}{2} \frac{L_{ab}(2) - L_{ba}(2)}{\epsilon_b - \epsilon_a}. \tag{6.35}$$

This choice suffers from numerical instability in the case $\epsilon_i \approx \epsilon_j$ or $\epsilon_a \approx \epsilon_b$, but, contrary to the case of Equation (6.24) and (6.25), it offers a way for computing $P^{(2)}_{ij}$ and $P^{(2)}_{ab}$ that doesn't require $O(N^5)$ operations. Moreover, in a parallel implementation, where usually the work load is achieved by distributing independent ij pairs, the usage of Equation (6.34) allows to drastically reduce the algorithmic complexity as well as avoiding the recomputation of MP2 amplitudes [229].

Due to the large size of the orbital Hessian matrix $\mathbf{A}$ ($ov \times ov$), the linear system of Equation (6.27) is commonly solved by iterative techniques [219, 233–235]. According to these methods, rather than calculating and storing the full $\mathbf{A}$, which is computationally inaccessible even for relatively small systems, at each iteration, the matrix-vector product $\sum_{ia} X_{ai} A_{aibj}$ is computed, with $\mathbf{X}$ being a trial solution. In this respect, it has to be noted that the orbital Hessian is made of a Coulomb part, first term in Equation (6.28), and an Exchange part, last two term in Equation (6.28) [230]. These two updates of the matrix-vector product, have thus to be computed consistently to the way the Coulomb and Exchange contributions to the Fock matrix are calculated during the SCF procedure. In the actual case, this means that the former is obtained within the GPW scheme and the latter via 4-index ERIs. Again, for efficiency reasons, the AO representation is preferred so that sparsity can be exploited.

This leads to the following matrix-vector update for the Coulomb part:

$$\begin{aligned}\sum_{ia} X_{ai}(ai|bj) &= \sum_{ai} X_{ai} \sum_{\mu\nu\lambda\sigma} C_{\mu a} C_{\nu i} (\mu\nu|\lambda\sigma) C_{\lambda b} C_{\sigma j} \\ &= \sum_{\mu\nu\lambda\sigma} \left[\sum_{ai} C_{\mu a} X_{ai} C_{\nu i} \right] (\mu\nu|\lambda\sigma) C_{\lambda b} C_{\sigma j} \\ &= \sum_{\mu\nu\lambda\sigma} Y_{\mu\nu} (\mu\nu|\lambda\sigma) C_{\lambda b} C_{\sigma j} \\ &= \sum_{\lambda\sigma} C_{\lambda b} C_{\sigma j} (\sum_{\mu\nu} Y_{\mu\nu} \phi_\mu \phi_\nu | \lambda\sigma) = \sum_{\lambda\sigma} C_{\lambda b} C_{\sigma j} (\rho^Y | \lambda\sigma) \\ &= \sum_{\lambda\sigma} C_{\lambda b} C_{\sigma j} \int \phi_\lambda(\vec{r}) \phi_\sigma(\vec{r}) v^Y(\vec{r}) d\vec{r} \end{aligned} \tag{6.36}$$

where v^Y is the electrostatic potential obtained from the $\rho^Y(\vec{r}) = \sum_{\mu\nu}^{\text{AO}} Y_{\mu\nu} \phi_\mu(\vec{r}) \phi_\nu(\vec{r})$ density and $Y_{\mu\nu}$ is the back transformed matrix associated to the actual trial solution X_{ai}. The required update for the Exchange part via 4-index ERI [236] reads:

$$\begin{aligned} -\sum_{ia} X_{ia} &[(ab|ij) + (aj|bi)] = \\ &= -\sum_{ai} X_{ai} \sum_{\mu\nu\lambda\sigma} C_{\mu a} C_{\nu i} [(\mu\lambda|\nu\sigma) + (\mu\sigma|\lambda\nu)] C_{\lambda b} C_{\sigma j} \\ &= -\sum_{\mu\nu\lambda\sigma} \left[\sum_{ai} C_{\mu a} X_{ai} C_{\nu i} \right] [(\mu\lambda|\nu\sigma) + (\mu\sigma|\lambda\nu)] C_{\lambda b} C_{\sigma j} \\ &= -\sum_{\mu\nu\lambda\sigma} Y_{\mu\nu} [(\mu\lambda|\nu\sigma) + (\mu\sigma|\lambda\nu)] C_{\lambda b} C_{\sigma j} \\ &= -\sum_{\mu\nu\lambda\sigma} C_{\lambda b} C_{\sigma j} (\mu\lambda|\nu\sigma) [Y_{\mu\nu} + Y_{\nu\mu}] \end{aligned} \tag{6.37}$$

where the permutation symmetry of the AO-ERIs has been exploited.

Finally, the MP2 correction to the energy-weighted density matrix $W_{pq}^{(2)}$, is calculated as follow: Occupied-occupied block:

$$W_{ij}^{(2)} = \frac{1}{2} \left\{ W_{ij}^{(2)}[I] + W_{ij}^{(2)}[II] + W_{ij}^{(2)}[III] \right\} \tag{6.38}$$

$$W_{ij}^{(2)}[I] = 2 \sum_{a}^{virt} \sum_{Q}^{\text{AUX}} (ja|Q) \Gamma_{ia}^{Q} = \sum_{\mu}^{\text{AO}} C_{\mu j} L_{\mu i}(1) \tag{6.39}$$

$$W_{ij}^{(2)}[II] = (\epsilon_i + \epsilon_j) P_{ij}^{(2)} \tag{6.40}$$

$$W_{ij}^{(2)}[III] = \sum_{pq}^{\text{MO}} P_{pq}^{(2)} A_{pqij} \tag{6.41}$$

Virtual-Virtual block:

$$W_{ab}^{(2)} = \frac{1}{2}\left\{W_{ab}^{(2)}[I] + W_{ab}^{(2)}[II]\right\} \tag{6.42}$$

$$W_{ab}^{(2)}[I] = 2\sum_{i}^{occ}\sum_{Q}^{\text{AUX}}(ib|Q)\Gamma_{ia}^{Q} = -\sum_{\nu}^{\text{AO}} C_{\nu b} L_{a\nu}(2) \tag{6.43}$$

$$W_{ab}^{(2)}[II] = (\epsilon_a + \epsilon_b) P_{ab}^{(2)} \tag{6.44}$$

Occupied-Virtual block:

$$W_{ai}^{(2)} = \frac{1}{2}\left\{W_{ai}^{(2)}[I] + W_{ai}^{(2)}[II]\right\} \tag{6.45}$$

$$W_{ai}^{(2)}[I] = 2\sum_{j}^{occ}\sum_{Q}^{\text{AUX}}(ji|Q)\Gamma_{ja}^{Q} = -\sum_{\nu}^{\text{AO}} C_{\nu i} L_{a\nu}(2) \tag{6.46}$$

$$W_{ai}^{(2)}[II] = 2\epsilon_i P_{ai}^{(2)} \tag{6.47}$$

The methodology presented up until here, is of general validity for any perturbation parameter x. In particular, for the calculation of the forces acting on the ions, the gradients of $E_{\text{RI}}^{(2)}$ with respect to the atomic positions have to be computed. Thus, within the GPW scheme, for which densities are represented in terms of both atom centered Gaussians and Plane Waves, only the derivatives of the former have to be considered since the latter are originless functions and do not depend of the atomic positions.

The RI-MP2 contribution to the total stress tensor is calculated according to [237–239]:

$$\Pi_{\alpha\beta}^{(2)} = -\frac{1}{3V}\sum_{\gamma=1}^{3}\frac{\partial E_{\text{RI}}^{(2)}}{\partial h_{\alpha\gamma}} h_{\gamma\beta}^{T} \tag{6.48}$$

where $h_{\alpha\gamma}$ are elements of the matrix of the cell vectors (Bravais lattice vectors) given by a_1, a_2 and a_3, that is $\mathbf{h} = [a_1, a_2, a_3]$. According to this, a change in $h_{\alpha\gamma}$ not only results in a scaling of all atom coordinates, but also affects the grid points over which the electrostatic densities and potentials are defined within the GPW scheme. The calculation of the integral derivatives in Equation (6.13) has thus to account for this effect, involving that additional terms have to be considered.

Again, the computation of $\mathbf{\Pi}^{(2)}$ can be split in two terms:

$$\mathbf{\Pi}^{(2)} = \mathbf{\Pi}^{(2)\text{-NS}} + \mathbf{\Pi}^{(2)\text{-S}} \tag{6.49}$$

that is, the non-separable part $\mathbf{\Pi}^{(2)\text{-NS}}$, specific of the RI-MP2 method, associated to the first two summations in Equation (6.13), and the separable part $\mathbf{\Pi}^{(2)\text{-S}}$, giving the additional contribution derived from the contraction of the MP2 relaxed density matrices with the stress derivative of the Fock and Overlap matrix elements, last summation in Equation (6.13).

Since in the non-separable part of $E_{\text{RI}}^{(2)^x}$, only integrals involving Coulomb interaction are required, $\mathbf{\Pi}^{(2)\text{-NS}}$ is obtained on the same line as done for the the stress tensor of the Hartree energy [176], for which the grid dependent contributions are evaluated following

the work of Corso and Resta [240]. This leads respectively for the 3- and 2-center contributions ($\Pi^{(2)\text{-NS}}_{\alpha\beta} = 4\Pi^{(2)\text{-NS-3c}}_{\alpha\beta} - 2\Pi^{(2)\text{-NS-2c}}_{\alpha\beta}$) to:

$$\begin{aligned}\Pi^{(2)\text{-NS-3c}}_{\alpha\beta} = -\frac{1}{3V}\Bigg[&\delta_{\alpha\beta}\left(\sum_{Q}^{\text{AUX}}\sum_{\mu\nu}^{\text{AO}}\Gamma^{Q}_{\mu\nu}(\mu\nu|Q)\right)\\ &+\sum_{Q}^{\text{AUX}}\sum_{\mu\nu}^{\text{AO}}\left[\Gamma^{Q}_{\mu\nu}+\Gamma^{Q}_{\nu\mu}\right]\int(R_{I\beta}-r_\beta)\nabla_{I\alpha}\phi_\mu(\vec{r})\phi_\nu(\vec{r})v^{Q}_{H}(\vec{r})d\vec{r}\\ &+\sum_{Q}^{\text{AUX}}\int(R_{I\beta}-r_\beta)\nabla_{I\alpha}\chi_Q(\vec{r})v_H^{\Gamma^Q}(\vec{r})d\vec{r}\\ &+\sum_{Q}^{\text{AUX}}\int\int\rho^{Q}(\vec{r})\rho^{\Gamma^Q}(\vec{r}')\frac{(r_\alpha-r'_\alpha)(r_\beta-r'_\beta)}{|\vec{r}-\vec{r}'|^3}d\vec{r}d\vec{r}'\Bigg]\end{aligned} \tag{6.50}$$

$$\begin{aligned}\Pi^{(2)\text{-NS-2c}}_{\alpha\beta} = -\frac{1}{3V}\Bigg[&\delta_{\alpha\beta}\left(\sum_{PQ}^{\text{AUX}}\Gamma^{PQ}(P|Q)\right)\\ &+2\sum_{Q}^{\text{AUX}}\int(R_{I\beta}-r_\beta)\nabla_{I\alpha}\chi_Q(\vec{r})v_H^{\Gamma^Q}(\vec{r})d\vec{r}\\ &+\sum_{Q}^{\text{AUX}}\int\int\rho^{Q}(\vec{r})\rho^{\Gamma^Q}(\vec{r}')\frac{(r_\alpha-r'_\alpha)(r_\beta-r'_\beta)}{|\vec{r}-\vec{r}'|^3}d\vec{r}d\vec{r}'\Bigg]\end{aligned} \tag{6.51}$$

where V is the volume of the cell, $\delta_{\alpha\beta}$ is the Kronecker delta, $\nabla_{I\alpha}$ is the α component of the gradient with respect to the atomic position and $R_{I\beta}$ refers to the β component of the atom coordinate. All other terms appearing in the above expressions have the same definitions given in Equation (6.18) and (6.19), note that the quantities labeled with the Γ^Q superscript are computed differently for $\Pi^{(2)\text{-NS-3c}}_{\alpha\beta}$ and $\Pi^{(2)\text{-NS-2c}}_{\alpha\beta}$. In both cases the first term arises from the scaling of the system's volume while the last is associated with the derivative of the electrostatic potential v_H [238]. The remaining components are associated to the derivatives of the Gaussian basis functions [241, 242].

The RI-MP2 stress tensor is completed with the separable part $\mathbf{\Pi}^{(2)\text{-S}}$. This final update is performed together with the calculation of the stress components of the Hartree-Fock energy. The approach is relatively straightforward and goes on the same line of the methodology previously explained for the general derivative case. Again, special care has to be considered in the case of the Hartree energy term, for which additional terms arise due to the dual representation of the density in GPW. These additional contributions are obtained employing a similar approach as done in the case of the non-separable part but starting from Equations (6.22) and (6.23).

In this section the general equations necessary for calculating the RI-MP2 energy derivatives have been presented with a particular focus on the way each term is calculated in the GPW framework. The presented approach has been applied for the calculation of the forces acting on the nuclei, and the stress tensor components. In summary, among all intermediates, only few quantities can be recognized as fundamental and need to be constructed in order to compute all the others, that is:

- $\Gamma^P_{ia}, \Gamma^{PQ} \rightarrow$ RI-MP2 non-separable correction to the 2-particle density matrix.
- $P^{(2)}_{ij}, P^{(2)}_{ab} \rightarrow$ Occupied-Occupied and Virtual-Virtual blocks of the MP2 correction to the 1-particle density matrix.
- $L_{\mu j}(1), L_{b\nu}(2) \rightarrow$ Occupied and Virtual Lagrangian in the mixed AO-MO representation.

6.3 Parallel Implementation

The general flowchart of the algorithm for the calculation of the RI-MP2 energy derivatives can be summarized as follow:

1. Calculation of $(P|Q)$ and its inverse square root $V^{-1/2}_{PQ}$, subsequent evaluation of B^P_{ia}. These intermediates are evaluated within the RI-GPW approach.
2. Formation of the $(ia|jb)_{\mathrm{RI}}$ integrals from B^P_{ia} (Equation (6.4)), calculation of $E_{\text{RI-MP2}}$, assembly of Y^P_{ia} and $P^{(2)}_{ab}$ according respectively with Equation (6.16) and (6.25), evaluation of $P^{(2)}_{ij}$ via Equation (6.24) for the diagonal and almost degenerate ($\epsilon_i \approx \epsilon_j$) elements.
3. Generation of $\Gamma^P_{ia}, \Gamma^{PQ}$ from Y^P_{ia}, B^P_{ia} and $V^{-1/2}_{PQ}$ (Equation (6.15) and (6.17)), evaluation of the non-separable contributions to $E^{(2)^x}_{\mathrm{RI}}$ and assembly of the $L_{\mu j}(1)$, $L_{b\nu}(2)$, occupied and virtual Lagrangian in the mixed AO-MO representation (Equation (6.32) and (6.33)).
4. Completion of $P^{(2)}_{ij}$ with $L_{ij}(1)$ for the non singular elements (Equation (6.34)). Construction of the RI-MP2 specific Lagrangian L_{bj} and solution of the Z-vector equations. Assembly of $P^{(2)}_{pq}$ and $W^{(2)}_{pq}$ and final evaluation of $E^{(2)^x}_{\mathrm{RI}}$ by contraction with $F^{(x)}_{pq}$ and $S^{(x)}_{pq}$.

Each of these step is described in the following subsections, with a particular focus on the parallelization strategy.

6.3.1 Evaluation of $V^{-1/2}_{PQ}$ and B^P_{ia} with the RI-GPW Method

The parallel algorithm for the evaluation of $V^{-1/2}_{PQ}$ and B^P_{ia} within the RI-GPW approach has been presented in great details in Ref. [131]. Here only the most important features of the parallel implementation are recalled in order to help the description of the algorithm in the next sections.

The parallelization strategy is based on a two level distribution of the workload, obtained by splitting the total N_p processes available into N_G groups, consisting of N_w processes ($N_p = N_G N_w$). The first level is associated to the work performed for a single given auxiliary basis function χ_P or vector $|P) = \sum_Q \chi_Q(\vec{r}) V^{-1/2}_{QP}$. The parallelization of at this level is obtained within the N_w processes of each group based on a hybrid OpenMP/MPI scheme involving, for example, parallel FFTs, halo-exchanges, numerical

integration of the basis functions over the electrostatic potential and sparse matrix multiplications. The second level corresponds to a distribution of these nearly independent calculations among the different groups. This is achieved by splitting the total number of auxiliary basis function N_a into N_G ranges $[P^{n_P}_{start}, P^{n_P}_{end}]$, each of them labeled with a given n_P coordinate, and assigned to the corresponding group. Additionally, each of the N_w processes within a group is given an index n_w, so that a processes is uniquely identified by its coordinate (n_P, n_w). Finally, the a virtual index is split in N_w ranges $[a^{n_w}_{start}, a^{n_w}_{end}]$, while a splitting of the occupied i is not necessary.

The workload distribution described so far allows for a scalable parallel implementation for the integral computation. In fact, the intense steps of communication are restricted within the group, made of a small number of tasks, while each group works independently for the different χ_P or $|P)$ associated to its range $[P^{n_P}_{start}, P^{n_P}_{end}]$.

Focusing on the calculation of B^P_{ia}, for each $P \in [P^{n_P}_{start}, P^{n_P}_{end}]$ the computational procedure can be summarized as follow:

- Evaluation of the density $\rho^P(\vec{r}) = \sum_Q \chi_Q(\vec{r}) V^{-1/2}_{QP}$ on the real space grid.
- Calculation of the electrostatic potential $v^P_H(\vec{r})$ associated to $\rho^P(\vec{r})$. This is obtained by first transferring $\rho^P(\vec{r})$ from the real to reciprocal space via FFT, solving the Poisson equation in Fourier space and finally back transferring the potential, with an additional FFT, from reciprocal to real space.
- Integration of the potential over the pairs of basis set elements, $B^P_{\mu\nu} = \int \phi_\mu(\vec{r})\phi_\nu(\vec{r}) v^P_H(\vec{r}) d\vec{r}$.
- Transformation of $B^P_{\mu\nu}$ from the AO to the MO basis by two consecutive matrix-matrix multiplication, that is $B^P_{i\nu} = \sum_\mu C_{\mu i} B^P_{\mu\nu}$ and finally $B^P_{ia} = \sum_\nu B^P_{i\nu} C_{\nu a}$.

The asymptotically dominating part of this procedure is associated to the last index transformation that has a formal scaling of $O(ovnN_a/N_p)$ while the integration of the potential has a cost that grows only quadratically with the system size. Nevertheless, due to the small prefactor associated to the former, the latter is usually more computationally demanding, even for relatively large systems [131].

At the end of this step each process stores the elements of B^P_{ia} for all i, $P \in [P^{n_P}_{start}, P^{n_P}_{end}]$ and $a \in [a^{n_w}_{start}, a^{n_w}_{end}]$.

6.3.2 Evaluation of the $O(N^5)$ Scaling Intermediates

With $O(N^5)$ scaling intermediates it is meant all the quantities that require for their construction a quintic computational effort. Within the RI-MP2 method these intermediates are $(ia|jb)_{\mathrm{RI}}$, Y^P_{ia}, $P^{(2)}_{ab}$ and $P^{(2)}_{ij}$ evaluated respectively with Equations (6.4), (6.16), (6.25) and (6.24) for which the formal computational effort grows as $O(o^2v^2N_a)$, $O(o^2v^2N_a)$, $O(o^2v^3)$ and $O(o^3v^2)$. The efficient construction of these intermediates is of prime importance since they are the asymptotically most expensive calculations of the RI-MP2 energy derivatives evaluation.

While in a serial algorithm their computation is relatively straightforward, a parallel implementation has to face the problems connected with the distributed storage of the precursors, such as B^P_{ia}, as well as the balancing of the workload over processes. Regarding the latter issue, a distribution of independent ij pairs ($i \leq j$) is of particular convenience

According to available memory define batch size B_S and size N_r of replication group (R)
Split the N_G groups into N_R subgroups ($N_R = N_G/N_r$)
Assign to each group G its coordinate in the subgroup n_R
Define the set $\{n_{R,w}\}$ including all processes having the same coordinate (n_R, n_w)
Define new set $\{P_{n_R,n_w}\}$ given as the union of $[P^{n'_P}_{start}, P^{n'_P}_{end}]$, $n'_P \in \{n_{R,w}\}$
Create β^P_{ia} by collecting B^P_{ia}, $P \in \{P_{n_R,n_w}\}$ from all other processes in $\{n_{R,w}\}$
Assign to each group G its set $\{IJ\}_G$ of IJ batches, ($I \leq J$, size of I and J given by B_S)
Loop over IJ batches ($IJ \in \{IJ\}_G$)
Collect $A^i_{aP} = \beta^P_{ia}$ and $E^j_{aP} = \beta^P_{ja}$, ($i \in I$, $j \in J$, all P, $a \in [a^{n_w}_{start}, a^{n_w}_{end}]$)
from all other processes in R with my same n_w coordinate $o^2 v N_a/(N_p B_S)$ (1)
Loop over ij ($i \in I$, $j \in J$)
Loop over n'_w, $n'_w \in G$
Collect E^j_{bP}, ($b \in [a^{n'_w}_{start}, a^{n'_w}_{end}]$)
$I_{ab} = (ia|jb) = \sum_P A^i_{aP} E^j_{bP}$, ($a \in [a^{n_w}_{start}, a^{n_w}_{end}], b \in [a^{n'_w}_{start}, a^{n'_w}_{end}]$) $o^2 v^2 N_a/N_p$ (2)
End n'_w Loop
$t_{ab} = (2I_{ab} - I_{ba})/\Delta^{ab}_{ij}$ $(*)$
$E^{(2)} = E^{(2)} + (2 - \delta_{ij}) \sum_{ab} I_{ab} t_{ab}$, ($a \in [a^{n_w}_{start}, a^{n_w}_{end}]$, all b) $o^2 v^2/N_p$ (3)
$P^{(2)}_{ab} = P^{(2)}_{ab} + \sum_c t_{ca} I_{cb}/\Delta^{cb}_{ij}$, ($a \in [a^{n_w}_{start}, a^{n_w}_{end}]$, all b) $(*)$ $o^2 v^3/N_p$ (4)
$P^{(2)}_{ii} = P^{(2)}_{ii} - \sum_{ab} t_{ab} I_{ab}/\Delta^{ab}_{ij}$, ($a \in [a^{n_w}_{start}, a^{n_w}_{end}]$, all b) $o^2 v^2/N_p$ (5)
$\Xi^i_{aP} = \Xi^i_{aP} + \sum_c t_{ac} E^j_{cP}$, ($a \in [a^{n_w}_{start}, a^{n_w}_{end}]$, all P, actual ij) $(*)$ $o^2 v^2 N_a/N_p$ (6)
if $i \neq j$ then
$P^{(2)}_{ab} = \sum_c t_{ac} I_{bc}/\Delta^{bc}_{ij}$, ($a \in [a^{n_w}_{start}, a^{n_w}_{end}]$, all b) $(*)$ $o^2 v^3/N_p$ (7)
$P^{(2)}_{jj} = P^{(2)}_{jj} - \sum_{ab} t_{ab} I_{ab}/\Delta^{ab}_{ij}$, ($a \in [a^{n_w}_{start}, a^{n_w}_{end}]$, all b) $o^2 v^2/N_p$ (8)
$\Lambda^j_{aP} = \Lambda^j_{aP} + \sum_c t_{ca} A^i_{cP}$, ($a \in [a^{n_w}_{start}, a^{n_w}_{end}]$, all P, actual ij) $(*)$ $o^2 v^2 N_a/N_p$ (9)
end if
End ij Loop
Collect Ξ^i_{aP} and Λ^j_{aP} from all other processes in R with my same n_w coord.: $o^2 v N_a/(N_p B_S)$ (10)
Accumulate $X^P_{ia} = X^P_{ia} + \Xi^i_{aP}$, if $i \neq j$ also $X^P_{ja} = X^P_{ja} + \Lambda^j_{aP}$ $o^2 v N_a/N_p$ (11)
(all ij processed in this cycle by all $G' \in R$, $P \in \{P_{n_R,n_w}\}$, $a \in [a^{n_w}_{start}, a^{n_w}_{end}]$)
End IJ Loop
Collect X^P_{ia} from all processes in $\{n_{R,w}\}$: $Y^P_{ia} = Y^P_{ia} + X^P_{ia}$, (all i, $P \in [P^{n_P}_{start}, P^{n_P}_{end}]$, $a \in [a^{n_w}_{start}, a^{n_w}_{end}]$)
Global summation of $P^{(2)}_{ab}$ across all process having the same coordinate n_w
Global summation of $E^{(2)}$ and diagonal elements $P^{(2)}_{ii}$ across all process

Figure 6.1: Pseudocode of the parallel algorithm for computing $E^{(2)}$, $P^{(2)}_{ab}$, Y^P_{ia} and the diagonal elements of $P^{(2)}_{ij}$. All statements labeled with $(*)$ involves in-group communication, these operations are shown explicitly in the algorithm only for the calculation of I_{ab}, in all other cases the procedure is performed likewise.

for the evaluation of $(ia|jb)_{\mathrm{RI}}$, Y^P_{ia} and $P^{(2)}_{ab}$, but not for the $P^{(2)}_{ij}$, for which the distribution of the ab pairs would be preferred. In order to overcome this complication, while retaining the ease of the ij distribution as well as avoiding additional events of communication, the $P^{(2)}_{ij}$ are evaluated via Equation (6.24) only for the diagonal and almost degenerate elements ($|\epsilon_i - \epsilon_j| < t_{\mathrm{sing}}$) while employing Equation (6.34) for all the other. This approach is referred as semi-canonical and t_{sing} is a threshold for discriminating which of the ij pairs have to be treated as almost degenerate, expected anyways to be a small fraction of

Compute the number d of ij pairs ($i < j$) for which $|\epsilon_i - \epsilon_j| < t_{\text{sing}}$
Define the set $\{ij\}_{\text{sing}}$ of size d containing the almost degenerate ij pairs
Assign to each group G its set $\{ijk\}_G$ of ijk triplets, ($ij \in \{ij\}_{\text{sing}}$, all occupied k)
Loop over ijk triplets ($ijk \in \{ijk\}_G$)
  Collect $A^i_{aP} = \beta^P_{ia}$, $E^j_{aP} = \beta^P_{ja}$ and $U^k_{aP} = \beta^P_{ka}$ (actual ijk, all P, $a \in [a^{n_w}_{start}, a^{n_w}_{end}]$)
    from all other processes in R with my same n_w coordinate $dovN_a/N_p$ (1)
  Loop over n'_w, $n'_w \in G$
    Collect U^k_{bP}, ($b \in [a^{n'_w}_{start}, a^{n'_w}_{end}]$)
    $I^{ik}_{ab} = (ia|kb) = \sum_P A^i_{aP} U^k_{bP}$, ($a \in [a^{n_w}_{start}, a^{n_w}_{end}], b \in [a^{n'_w}_{start}, a^{n'_w}_{end}]$) dov^2N_a/N_p (2)
    $I^{jk}_{ab} = (ja|kb) = \sum_P E^j_{aP} U^k_{bP}$, ($a \in [a^{n_w}_{start}, a^{n_w}_{end}], b \in [a^{n_w}_{start}, a^{n_w}_{end}]$) dov^2N_a/N_p (3)
  End n'_w Loop
  $t^{ik}_{ab} = (2I^{ik}_{ab} - I^{ik}_{ba})/\Delta^{ab}_{ik}$ $(*)$
  $P_{\text{val}} = -\sum_{ab} t^{ik}_{ab} I^{jk}_{ab} / \Delta^{ab}_{jk}$, ($a \in [a^{n_w}_{start}, a^{n_w}_{end}]$, all b) dov^2/N_p (4)
  $P^{(2)}_{ij} = P^{(2)}_{ij} + P_{\text{val}}$
  $P^{(2)}_{ji} = P^{(2)}_{ji} + P_{\text{val}}$
End ijk Loop
Global summation of $P^{(2)}_{ij}$ ($ij \in \{ij\}_{\text{sing}}$) across all process

Figure 6.2: Pseudocode of the parallel algorithm for computing the off diagonal elements of $P^{(2)}_{ij}$ for almost degenerate ij pairs ($\epsilon_i \approx \epsilon_j$). The meaning of the symbols is the same as in Figure 6.1.

the total.

The pseudocode for the parallel evaluation of $(ia|jb)_{\text{RI}}$, Y^P_{ia}, $P^{(2)}_{ab}$ and the diagonal elements $P^{(2)}_{ii}$ is sketched in Figure 6.1, while the update of $P^{(2)}_{ij}$ for the almost degenerate pairs is shown in Figure 6.2. The completion of $P^{(2)}_{ij}$ for the remaining elements is performed later, when $L_{\mu j}(1)$ is made available. In the pseudocodes the expected computational effort, expressed in term of order of, is reported for the most important operations.

The algorithm shown in Figure 6.1 closely resembles the structure employed for the RI-MP2 energy evaluation described in Ref. [131]. As mentioned in the previous section the B^P_{ia} intermediate is distributed such that each process stores elements for all occupied i, $P \in [P^{n_P}_{start}, P^{n_P}_{end}]$ and $a \in [a^{n_w}_{start}, a^{n_w}_{end}]$. That is, the virtual index a is distributed over a small number of MPI tasks within the group G while the auxiliary index P is distributed over the large amount of N_G groups. The same kind of distribution has been adopted for Y^P_{ia} while $P^{(2)}_{ab}$ and $P^{(2)}_{ij}$ are created in a replicated form within each group, only the virtual index a of the former is distributed over the elements of the group with the usual $a \in [a^{n_w}_{start}, a^{n_w}_{end}]$.

The parallelization is achieved by distributing independent ij pairs over the N_G groups. Each group, for a given ij pair, perform the following set of operations:

1. B^P_{ia} is redistributed such that the full range of the auxiliary index P, for i and j, is collected on local buffers from all other groups, while keeping the virtual index distribution, $a \in [a^{n_w}_{start}, a^{n_w}_{end}]$

2. the $(ia|jb)_{\text{RI}}$ integrals are generated for the actual ij pair in a matrix-multiplication fashion (Equation (6.4))

3. the amplitudes t_{ij}^{ab} are formed from $(ia|jb)_{\mathrm{RI}}$ and $\Delta_{ij}^{ab} = \epsilon_i + \epsilon_j - \epsilon_a - \epsilon_b$

4. the contributions to $P_{ab}^{(2)}$ and $P_{ii}^{(2)}$ are accumulated into the relative local buffers

5. the contribution to Y_{ia}^{P} are accumulated into two intermediates namely Ξ_{aP}^{i} and Λ_{aP}^{j} distributed within the group such that each process stores the full auxiliary index P and $a \in [a_{start}^{n_w}, a_{end}^{n_w}]$

6. Ξ_{aP}^{i} and Λ_{aP}^{j} are redistributed over all groups and accumulated into the local Y_{ia}^{P}.

In the above procedure, the first and the last steps involves inter-group communication. They can be considered isomorphic with the difference that in the former, for the actual ij pair, each process collect the full range of auxiliary functions for $a \in [a_{start}^{n_w}, a_{end}^{n_w}]$, in the latter, each task collects all the ij indices processed by all other groups for its preassigned range of $P \in [P_{start}^{n_P}, P_{end}^{n_P}]$ and $a \in [a_{start}^{n_w}, a_{end}^{n_w}]$. All other operations are performed locally within the group, that is, only a small amount of communication takes place restricted to the members of the group.

The main source of communication of the parallel algorithm is thus related to the inter-group redistribution steps mentioned above. According to the implementation designed for the RI-MP2 energy, three key aspects are considered in order to enhance the efficiency of these operations:

- B_{ia}^{P} is replicated into β_{ia}^{P} distributed within subgroups (the replication groups R). This allows to restrict the number of processes that have to communicate at each cycle.

- The messages are exchanged employing point-to-point communication. This allows to avoid global synchronization of processes while keeping a low memory usage.

- The ij pairs are communicated in batches, reducing the overall number of messages.

While the first point is more a technical ingredient, the other two are mandatory in order to obtain a scalable implementation, so that the required communication is reduced increasing the number of processes.

In more details, following the pseudocode given in Figure 6.1, as a first stage, according to the available memory the size N_r of the replication groups R is defined and the elements of B_{ia}^{P} are replicated into β_{ia}^{P} distributed among the members of R. N_R being the number of replication groups, a ratio N_R/N_G of ~ 0.1, has been shown to be a good compromise between the time necessary for the replication and the gain in communication in the subsequent steps. The remaining memory available per process is then used to define the maximum batch size B_S and the total number of IJ batches ($I \leq J$) are then distributed statically over the N_G groups. To achieve the best possible load balance, the number of IJ batches is restricted to be a multiple of the number of groups N_G and the remaining ij single pairs are again statically distributed over groups.

At this point each group loops over its preassigned IJ batches and, as a first task, collect from all other members of the replication group R, the elements $A_{aP}^{i} = \beta_{ia}^{P}$ and $E_{aP}^{j} = \beta_{ja}^{P}$. For all ij pairs in the actual IJ batch the operations between 2 and 5 previously described are performed. This allows to calculate the contribution coming from the actual ij to $P_{ab}^{(2)}$,$P_{ii}^{(2)}$,$P_{jj}^{(2)}$ and Ξ_{aP}^{i}, Λ_{aP}^{j}. Before moving to the next batch, the

second inter-group communication step take place, redistributing Ξ^i_{aP} and Λ^j_{aP} within the members of the replication group. This corresponds in the pseudocode to line (10) and (11), for which the intermediate X^P_{ia} is introduced. This quantity stays to Y^P_{ia} as β^P_{ja} stays to B^P_{ja}, *i.e.* X^P_{ia} represents the replicated version of Y^P_{ia} collecting the contributions coming from all the ij pairs processed by the groups G in the replication group R. For this reason, at the end of the loop over IJ batches, an additional step of decomposition of X^P_{ia} is required in order to generate Y^P_{ia} in its final form, that is, distributed such that each process stores the elements for all occupied i, $P \in [P^{n_P}_{start}, P^{n_P}_{end}]$ and $a \in [a^{n_w}_{start}, a^{n_w}_{end}]$. The two final steps are global summation of the elements of $P^{(2)}_{ab}$ and diagonal of $P^{(2)}_{ij}$ with the difference that the latter is summed over all processes while the former only across those that share the same virtual index range $a \in [a^{n_w}_{start}, a^{n_w}_{end}]$, *i.e* those labeled with the same coordinate n_w.

At this stage, what remains to be done is the update $P^{(2)}_{ij}$ with the potentially singular ij pairs. This is accomplished first by checking the total number d of almost degenerate ij pairs, *i.e.* the pairs for which $|\epsilon_i - \epsilon_j| < t_{\text{sing}}$. For each of these a loop over all occupied k is performed, the $(ia|kb)_{\text{RI}}$ and $(ja|kb)_{\text{RI}}$ are generated, and the $P^{(2)}_{ij}$ element is updated according to Equation (6.24).

As shown by the pseudocode in Figure 6.2, the parallelization of these steps is obtained in a very similar way as done for the computation of $P^{(2)}_{ab}$, with the main difference that in this case the ijk triplets are distributed over the N_G groups. The number of ijk triplets $(d \times o)$ is usually small compared to the total number of ij pairs (o^2), for this reason, the communication scheme employing batches is not exploited since it may lead to poor balance of the work load. As in the previous case, the procedure is finalized with a global summation over all processes of the almost degenerate elements of $P^{(2)}_{ij}$. It has to be noted that, in the case that no potentially singular ij pairs are detected (as in most of the cases), this part of the algorithm is completely skipped.

As a summary, the parallel algorithm described in this section can be splitted into two relevant parts: communication and computation of the $O(N^5)$ scaling intermediates. The first part has a cost that can be estimated to be $O(o^2 v N_a/(N_p B_S))$, that is derived by considering that the total number of messages exchanged by each process is $O(o^2/(B_S^2 N_G))$ while the time required for each event of communication (considered to be proportional to the message sizes) is $O(v N_a B_S/N_w)$. This implies that communication is eventually an $O(N^4)$ operation whose effort scales with the number of processes $N_p = N_w N_G$. It has to be noted that, compared to the energy RI-MP2 algorithm, this operation is expected to be roughly two times more expensive since it involves not only the redistribution of β^P_{ia}, line (1) in the pseudocode 6.1, but also the similar operation for Ξ^i_{aP} and Λ^j_{aP}, line (10). Concerning the computation of the $O(N^5)$ intermediates, these are reported in the pseudocode 6.1 at line (2), the generation of the $(ia|jb)_{\text{RI}}$ integrals, at line (4) and (7), update of $P^{(2)}_{ab}$ and at line (6) and (9), update of Ξ^i_{aP} and Λ^j_{aP}. Again, a comparison with the energy RI-MP2 algorithm, for which only the generation of the $(ia|jb)_{\text{RI}}$ is required, leads to the conclusion that, for the actual implementation, the $O(N^5)$ part is expected to be roughly 3 to 4 times more expensive. As an additional remark, all these tasks are accomplished as matrix multiplications, and thus the performance of highly optimized routines, such as DGEMM, can be exploited as well as accelerated by employing a hybrid implementation that utilizes graphics processing units (GPUs).

Redistribute Y_{ia}^P and B_{ia}^P into $(ov \times N_a)$ parallel distributed matrices $Y_{ia,P}$, $B_{ia,P}$
Create $(ov \times N_a)$ parallel distributed matrix $\Gamma_{ia,P}$
Create $(N_a \times N_a)$ parallel distributed matrices $\gamma_{P,Q}$ and $\Gamma_{P,Q}$
Perform parallel matrix-matrix multiplication:

$\Gamma_{ia,P} = \sum_Q Y_{ia,Q} V_{QP}^{-1/2}$ ovN_a^2/N_p (1)

$\gamma_{P,Q} = \sum_{ia} \Gamma_{ia,P} B_{ia,Q}$ ovN_a^2/N_p (2)

$\Gamma_{P,Q} = \sum_R \gamma_{P,R} V_{RQ}^{-1/2}$ N_a^3/N_p (3)

Redistribute $\Gamma_{P,Q} \rightarrow \Gamma^{PQ}$, (all P, $Q \in [P_{start}^{n_P}, P_{end}^{n_P}]$)
Redistribute $\Gamma_{ia,P} \rightarrow \Gamma_{ia}^P$, ($P \in [P_{start}^{n_P}, P_{end}^{n_P}]$, for each group all i, a
stored in the form of a parallel distributed matrix defined within the group)
Create $L_{\mu i}^G(1)$ and $L_{a\nu}^G(2)$ parallel distributed matrices defined within the group
Loop over P auxiliary basis functions ($P_{start}^{n_P} \leq P \leq P_{end}^{n_P}$)

Calculate density $\rho^P(\vec{r}) = \chi_P(\vec{r})$ on the real space grid $N_a S/N_p$ (4)

Transfer $\rho^P(\vec{r}) \rightarrow \rho^P(\vec{G})$: $\rho^P(\vec{G}) = \mathbf{FFT}[\rho^P(\vec{r})]$ $N_a S \log(S)/N_p$ (5)

Solve Poisson's Equation: $\rho^P(\vec{G}) \rightarrow v_H^P(\vec{G})$ $N_a S \log(S)/N_p$ (6)

Transfer $v_H^P(\vec{G}) \rightarrow v_H^P(\vec{r})$: $v_H^P(\vec{r}) = \mathbf{FFT^{-1}}[v_H^P(\vec{G})]$ $N_a S \log(S)/N_p$ (7)

Calculate integral derivatives: $D_Q^P = 2 \int \chi_Q^x(\vec{r}) v_H^P(\vec{r}) d\vec{r}$ (all AUX Q) N_a^2/N_p (8)

$E_{\mathrm{RI}}^{(2)^x} = E_{\mathrm{RI}}^{(2)^x} - 2 \sum_Q \Gamma^{PQ} D_Q^P$, (all Q, actual P)

Back-transform Γ_{ia}^P, virtual index: $\Gamma_{i\nu}^P = \sum_a \Gamma_{ia}^P C_{\nu a}$ $ovnN_a/N_p$ (9)

Back-transform $\Gamma_{i\nu}^P$, occupied index and symmetrize: $\Gamma_{\mu\nu}^P = \sum_i C_{\mu i} \Gamma_{i\nu}^P$ $on^2 N_a/N_p$ (10)

Calculate integrals and derivatives:

$I_{\mu\nu}^P = \int \phi_\mu(\vec{r}) \phi_\nu v_H^P(\vec{r}) d\vec{r}$ (all $\mu\nu$) nN_a/N_p (11)

$D_{\mu\nu}^P = 2 \int \phi_\mu^x(\vec{r}) \phi_\nu v_H^P(\vec{r}) d\vec{r}$ (all $\mu\nu$) nN_a/N_p (12)

$E_{\mathrm{RI}}^{(2)^x} = E_{\mathrm{RI}}^{(2)^x} + 4 \sum_{\mu\nu} \Gamma_{\mu\nu}^P D_{\mu\nu}^P$, (all $\mu\nu$, actual P)

Transform first index of $I_{\mu\nu}^P$ to occupied MO: $I_{i\nu}^P = \sum_\mu C_{\mu i} I_{\mu\nu}^P$ onN_a/N_p (13)

Accumulate contribution to the Lagrangian in the mixed AO-MO representation:

$L_{\mu i}^G(1) = L_{\mu i}^G(1) + 2 \sum_\nu I_{\mu\nu}^P \Gamma_{i\nu}^P$ onN_a/N_p (14)

$L_{a\nu}^G(2) = L_{a\nu}^G(2) - 2 \sum_i I_{i\nu}^P \Gamma_{ia}^P$ $ovnN_a/N_p$ (15)

Calculate density $\rho^{\Gamma^P}(\vec{r}) = \sum_{\mu\nu} \Gamma_{\mu\nu}^P \phi_\mu(\vec{r}) \phi_\nu(\vec{r})$ on the real space grid $N_a S/N_p$ (16)

Transfer $\rho^{\Gamma^P}(\vec{r}) \rightarrow \rho^{\Gamma^P}(\vec{G})$: $\rho^{\Gamma^P}(\vec{G}) = \mathbf{FFT}[\rho^{\Gamma^P}(\vec{r})]$ $N_a S \log(S)/N_p$ (17)

Solve Poisson's Equation: $\rho^{\Gamma^P}(\vec{G}) \rightarrow v_H^{\Gamma^P}(\vec{G})$ $N_a S \log(S)/N_p$ (18)

Transfer $v_H^{\Gamma^P}(\vec{G}) \rightarrow v_H^{\Gamma^P}(\vec{r})$: $v_H^{\Gamma^P}(\vec{r}) = \mathbf{FFT^{-1}}[v_H^{\Gamma^P}(\vec{G})]$ $N_a S \log(S)/N_p$ (19)

$E_{\mathrm{RI}}^{(2)^x} = E_{\mathrm{RI}}^{(2)^x} + 4 \int \chi_P^x(\vec{r}) v_H^{\Gamma^P}(\vec{r}) d\vec{r}$ (actual P) N_a/N_p (20)

End P Loop
Create $L_{\mu i}(1)$ and $L_{a\nu}(2)$ parallel distributed matrices defined aver all processes
Redistribute $L_{\mu i}^G(1), L_{a\nu}^G(2) \rightarrow L_{\mu i}(1), L_{a\nu}(2)$ accumulating contributions coming from other processes

Figure 6.3: Pseudocode of the parallel algorithm for computing the mixed AO-MO Lagrangian $L_{\mu i}(1)$, $L_{a\nu}(2)$ and the non-separable contribution to $E_{\mathrm{RI}}^{(2)^x}$ with the GPW approach.

6.3.3 Non-Separable Contribution to $E_{\mathbf{RI}}^{(2)^x}$, Assembly of $L_{\mu j}(1)$ and $L_{b\nu}(2)$

The calculation of the non-separable contribution to $E_{\mathrm{RI}}^{(2)^x}$, $L_{\mu j}(1)$ and $L_{b\nu}(2)$ is performed within the same procedure since many intermediates, deriving for example from the back-transformation of Γ_{ia}^P, are in common between these evaluations and can

be computed within a loop over the auxiliary index P. In general, the non-separable contributions to $E_{\mathrm{RI}}^{(2)^x}$ are calculated by contraction of 3- and 2-center RI integral derivative with the non-separable correction to the 2-particle density matrix, computed in the framework of the RI-GPW approach (Equations (6.18) and (6.19)); while $L_{\mu j}(1)$ and $L_{b\nu}(2)$ are evaluated with Equation (6.32) and (6.33). The pseudocode of this procedure is reported in Figure 6.3.

Prior the calculation of these quantities it is thus necessary to assembly Γ_{ia}^{P} and Γ^{PQ} (3- and 2-index non-separable correction to the 2-PDM). This is achieved from the B_{ia}^{P}, Y_{ia}^{P} and $V_{PQ}^{-1/2}$ intermediates by a sequence of parallel matrix multiplications (lines (1), (2) and (3) in Figure 6.3). Since both B_{ia}^{P} and Y_{ia}^{P} are distributed such that each process stores the elements for all i, $P \in [P_{start}^{n_P}, P_{end}^{n_P}]$ and $a \in [a_{start}^{n_w}, a_{end}^{n_w}]$, a redistribution step is require in order to reorganize the data into the form of a parallel distributed matrix suitable for performing efficiently the multiplications. A second redistribution step is then performed for the computed Γ_{ia}^{P} and Γ^{QP} such that each group G stores for $P \in [P_{start}^{n_P}, P_{end}^{n_P}]$ all ia of the former, and all Q of the latter. The elements of Γ_{ia}^{P} reshuffled in this way are further organized within the group in the form of a parallel distributed matrix.

At this point the actual computation of the non-separable part of $E_{\mathrm{RI}}^{(2)^x}$ as well as $L_{\mu j}(1)$ and $L_{b\nu}(2)$ is performed by accumulating the contributions to these quantities associated to each auxiliary element P. As done for the calculation of B_{ia}^{P}, the parallelization is achieved by letting each group G work on its preassigned range of $P \in [P_{start}^{n_P}, P_{end}^{n_P}]$. Again all operations associated to a given P are performed in parallel within the members of the groups

The series of required operations for each P are shown in details in the pseudocode reported in Figure 6.3. As a first step, according to the RI-GPW scheme, the electrostatic potential $v_H^P(\vec{r})$ associated to the single auxiliary basis function $\chi_P(\vec{r})$ is evaluated and made available on the real space grid. The potential is thus integrated over the auxiliary basis function derivatives $\chi_Q^x(\vec{r})$ for all Q and subsequently contracted with the relative elements of Γ^{QP} giving the non-separable contribution to $E_{\mathrm{RI}}^{(2)^x}$ from the 2-center ERI's.

At this point two steps of back-transformation of Γ_{ia}^{P} are performed obtaining both $\Gamma_{i\nu}^{P}$ and $\Gamma_{\mu\nu}^{P}$. The previously calculated potential $v_H^P(\vec{r})$ is now integrated over the pair of primary basis functions $\phi_\mu(\vec{r})\phi_\nu(\vec{r}) \rightarrow I_{\mu\nu}^{P}$ and associated derivatives $\phi_\mu^x(\vec{r})\phi_\nu(\vec{r}) \rightarrow D_{\mu\nu}^{P}$. The integral derivatives $D_{\mu\nu}^{P}$ are contracted with the fully back-transformed $\Gamma_{\mu\nu}^{P}$ giving the first non-separable contribution to $E_{\mathrm{RI}}^{(2)^x}$ from the 3-center ERI's, while the plain integrals $I_{\mu\nu}^{P}$ are multiplied with $\Gamma_{i\nu}^{P}$ and accumulated into a local buffer $L_{\mu j}^{G}(1)$ of $L_{\mu j}(1)$. This update can exploit the sparsity of $I_{\mu\nu}^{P}$ making this step $O(on)$ for each P. On the other hand, the contribution to $L_{b\nu}(2)$ is calculated by first transforming the first index of $I_{\mu\nu}^{P}$ to the occupied MO $I_{i\nu}^{P}$ and then performing the update with Γ_{ia}^{P}, again obtained as a matrix multiplication and accumulate on the relative local buffer $L_{b\nu}^{G}(2)$. In this case, the matrices are not sparse, and thus the associated cost is $O(ovn)$ for each P. Finally, the second non-separable contribution to $E_{\mathrm{RI}}^{(2)^x}$ from the 3-center ERI's is computed by integrating the potential $v_H^{\Gamma^P}(\vec{r})$ associated to the $\sum_{\mu\nu}\Gamma_{\mu\nu}^{P}\phi_\mu(\vec{r})\phi_\nu(\vec{r})$ electrostatic density with the auxiliary basis function derivative $\chi_P^x(\vec{r})$ only for the actual P.

At the end of the loop over the auxiliary index P each group stores the two buffers $L_{\mu j}^{G}(1)$ and $L_{b\nu}^{G}(2)$ containing the contribution to $L_{\mu j}(1)$ and $L_{b\nu}(2)$ associated to the $P \in [P_{start}^{n_P}, P_{end}^{n_P}]$. In order to obtain $L_{\mu j}(1)$ and $L_{b\nu}(2)$ in their final form, *i.e.* defined

over all process, a redistribution step is required, for which each process receives and accumulates from all others the data associated with its new local portion of the two matrices.

Similarly to the case of the computation of B_{ia}^{P}, the asymptotically dominating steps of this procedure scales as $O(N^4)$. As shown in the pseudocode of Figure 6.3, these are associated to the calculation of Γ_{ia}^{P} and Γ^{PQ}, update of $L_{\mu j}(1)$ and $L_{b\nu}(2)$, and indices transformations AO↔MO. These operations display a relatively small prefactor since are basically performed as matrix multiplications. On the other hand the calculation of the integrals and their derivatives is linear scaling for each P since only pairs of overlapping Gaussians need to be considered, and only a finite number of grid points within a spherical region around the center of the primitive Gaussian functions is required. This makes the overall effort in the integral computation $O(N^2)$. Nevertheless this operation displays a quite large prefactor and results to take a large amount of the total time (30-40 %) even for relatively large systems.

This part of the algorithm is specific to the computation of the RI-MP2 energy derivatives, that means that it constitutes an overhead that is not necessary in the case for which only the energy is required. Even if the structure of the described procedure is similar to that employed for the evaluation of B_{ia}^{P}, in this case not only the integrals are computed, but also their derivatives. This implies a cost for integral computation that is roughly double that associated to the relative operation in the calculation of B_{ia}^{P}.

In the case the stress tensor has to be computed, additional operations have to be considered. These operations are not reported in Figure 6.3, but they can be derived by inspection of Equations (6.50) and (6.51). Note that in this case also elements of the type $(R_{I\beta} - r_\beta)\nabla_{I\alpha}\phi_\mu(\vec{r})\phi_\nu(\vec{r})$ have to be integrated, resulting in an additional overhead roughly equivalent to the computation of the integral derivatives.

6.3.4 Final evaluation of $E_{\mathbf{RI}}^{(2)^x}$

The evaluation of the RI-MP2 energy derivatives is completed by a series of operations that allow to generate the $P_{pq}^{(2)}$ and $W_{pq}^{(2)}$ in their final form. Once $P_{pq}^{(2)}$ and $W_{pq}^{(2)}$ are made available, their contraction with the skeleton derivatives of the Fock and Overlap matrix elements is performed at the same time with the evaluation of the derivatives of the HF energy leading to the final result.

The sequence of these operations is summarized in the pseudocode of Figure 6.4. The virtual-virtual block of $P_{pq}^{(2)}$ is already available from the procedure described in Section 6.3.2, while the occupied-occupied part has to be completed for the non-singular elements according to Equation (6.34). The occupied-virtual block is instead obtained as the solution of the Z-vector Equations (6.27).

In order to do so, first the RI-MP2 specific Lagrangian L_{bj} has to be assembled. As shown in Equation (6.29), to L_{bj} contributes four terms. The first two are calculated from $L_{\mu j}(1)$ and $L_{b\nu}(2)$ just by transforming the indices from the AO to the MO basis. The remaining two are computed by contraction of the the virt-virt and occ-occ blocks of $P_{pq}^{(2)}$ with the integrals generated by coupled-perturbed Hartree-Fock (CPHF) theory (the A_{pqrs} matrix). This contraction is often referred as CPHF-like update, and has a computational cost that is isomorphic to the update of the Fock matrix in the standard SCF procedure. Equations (6.36) and (6.37) display the operations that have to be performed for each

Create $P_{pq}^{(2)}$ and $W_{pq}^{(2)}$ parallel distributed matrices defined aver all processes
Fill virtual-virtual block $P_{ab}^{(2)}$, (all a, b)
Fill occupied-occupied block $P_{ij}^{(2)}$ (diagonal elements and $ij \in \{ij\}_{\text{sing}}$)
Transform first index of $L_{\mu i}(1)$ to occupied MO: $L_{ij}(1) = \sum_\mu C_{\mu j} L_{\mu i}(1)$
Complete occupied-occupied block: $P_{ij}^{(2)} = \frac{1}{2} \frac{L_{ij}(1) - L_{ji}(1)}{\epsilon_j - \epsilon_i}$, $(ij \notin \{ij\}_{\text{sing}})$
Calculate contributions to $W_{pq}^{(2)}$ from $P_{ij}^{(2)}$, $P_{ab}^{(2)}$, $L_{\mu i}(1)$ and $L_{a\nu}(2)$:

$$W_{ij}^{(2)} = \tfrac{1}{2} \left[\sum_\mu C_{\mu j} L_{\mu i}(1) + (\epsilon_i + \epsilon_j) P_{ij}^{(2)} \right]$$
$$W_{ab}^{(2)} = \tfrac{1}{2} \left[-\sum_\nu C_{\nu b} L_{a\nu}(2) + (\epsilon_a + \epsilon_b) P_{ab}^{(2)} \right]$$
$$W_{ai}^{(2)} = -\tfrac{1}{2} \sum_\nu C_{\nu i} L_{a\nu}(2)$$

Calculate Lagrangian L_{bj}: $L_{bj} = \sum_\mu C_{\mu b} L_{\mu j}(1) + \sum_\nu L_{b\nu}(2) C_{\nu j} + \sum_{ac} P_{ac}^{(2)} A_{acbj} + \sum_{ik} P_{ik}^{(2)} A_{ikbj}$
Calculate $P_{ai}^{(2)}$ by solving the Z-vector equations: $\sum_{ai} \left[\delta_{ij} \delta_{ab} (\epsilon_a - \epsilon_i) + A_{aibj} \right] P_{ai}^{(2)} = -L_{bj}$
Complete $P_{pq}^{(2)}$ with occupied-virtual block $P_{ai}^{(2)}$
Complete $W_{pq}^{(2)}$:

$$W_{ij}^{(2)} = W_{ij}^{(2)} + \tfrac{1}{2} \sum_{pq} P_{pq}^{(2)} A_{pqij}$$
$$W_{ai}^{(2)} = W_{ai}^{(2)} + \epsilon_i P_{ai}^{(2)}$$

Transform $P_{pq}^{(2)} \rightarrow P_{\mu\nu}^{(2)}$, $W_{pq}^{(2)} \rightarrow W_{\mu\nu}^{(2)}$ and symmetrize
Final ${E_{\text{RI}}^{(2)}}^x$ evaluation by contraction of $P_{\mu\nu}^{(2)}$ and $W_{\mu\nu}^{(2)}$ with $F_{\mu\nu}^x$ and $S_{\mu\nu}^x$

Figure 6.4: Pseudocode of the parallel algorithm for computing $P_{pq}^{(2)}$, $W_{pq}^{(2)}$ and the final contributions to ${E_{\text{RI}}^{(2)}}^x$.

CPHF-like update, note that in the actual case the summation ranges are over virt-virt and occ-occ orbitals respectively for the contraction with $P_{ab}^{(2)}$ and $P_{ij}^{(2)}$.

Once L_{bj} is assembled the Z-vector equations are solved employing the Pople method [219]. From a computational standpoint this is equivalent to solving a large system of linear equations with an iterative technique, for which, at each iteration, only the matrix-vector product (CPHF-like update) has to be performed. The parallelization of the CPHF-like update closely follows the scheme employed in CP2K for evaluation of the Fock matrix elements [91, 105] and will not be described further here.

Note that, for dense systems with large basis, the computation of the 4-index ERI's over AO, necessary to calculate the exchange part of the CPHF-like update, is by far the most demanding task of this procedure. Since these integrals are the same as those employed in the SCF procedure, if enough memory is available, they can be stored in core, avoiding their recomputation. This can greatly speed-up the solution of the Z-vector equations.

The RI-MP2 correction to the energy-weighted density matrix $W_{pq}^{(2)}$ is finally generated from $P_{pq}^{(2)}$, $L_{\mu j}(1)$ and $L_{b\nu}(2)$ according to Equations (6.38)-(6.47), for which an additional CPHF-like update is required for the occupied-occupied block. With these matrices defined, previous a step of back-transformation from the AO to the MO basis, the derivatives of the total energy (RI-MP2 + HF) can be finalized by contraction with the skeleton derivatives of the Fock $F_{\mu\nu}^{(x)}$ and Overlap $S_{\mu\nu}^{(x)}$ matrix elements.

Table 6.1: Memory usage in the different part of the parallel algorithm expressed as "order of" the calculation parameters. n and N_a number of primary and auxiliary basis functions, o and v number of occupied and virtual orbitals, S grid size, N_G and N_w number of groups and group size, N_R and N_r number and size of the replication group ($N_G = N_R N_r$), B_S batch size for ij pairs. N_p number of processes. N_G, N_w and N_p are related by $N_p = N_G N_w$. The notation employed for the entries is referred to the different algorithms.

	Memory
Evaluation of $V_{PQ}^{-1/2}$ and B_{ia}^{P} (Section 6.3.1):	
$\rho(\vec{r})$, $v(\vec{r})$, $\rho(\vec{G})$, $v(\vec{G})$	S/N_w
$V_{QP}^{-1/2}$ - B_{ia}^{P}	N_a^2/N_p - ovN_a/N_p
$O(N^5)$ Scaling Intermediates (Section 6.3.2):	
β_{ia}^{P}, X_{ia}^{P} - Y_{ia}^{P}	$ovN_a/(N_r N_w)$ - ovN_a/N_p
A_{aP}^{i}, E_{aP}^{j}, Ξ_{aP}^{i}, Λ_{aP}^{j}	vN_aB_S/N_w
I_{ab}, t_{ab}, $P_{ab}^{(2)}$ - $P_{ij}^{(2)}$	v^2/N_w - o^2
Non-Separable $E_{\mathrm{RI}}^{(2)^x}$, $L_{\mu j}(1)$ and $L_{b\nu}(2)$ (Section 6.3.3)	
$\rho(\vec{r})$, $v(\vec{r})$, $\rho(\vec{G})$, $v(\vec{G})$	S/N_w
Γ^{PQ} - Γ_{ia}^{P}	N_a^2/N_p - ovN_a/N_p
$\Gamma_{\mu\nu}^{P}$, $I_{\mu\nu}^{P}$ - $\Gamma_{i\nu}^{P}$, $I_{i\nu}^{P}$, $L_{\mu j}^{G}(1)$ - $L_{b\nu}^{G}(2)$	n/N_w - on/N_w - vn/N_w
$L_{\mu j}(1)$ - $L_{b\nu}(2)$	on/N_p - vn/N_p
Final evaluation of $E_{\mathrm{RI}}^{(2)^x}$ (Section 6.3.4)	
$P_{pq}^{(2)}$, $W_{pq}^{(2)}$ - $P_{ia}^{(2)}$, L_{jb}	n^2/N_p - ov/N_p
CPHF-like update	$N_{\text{AO-ERI}}/N_p$

6.3.5 Memory Usage

The RI-MP2 method, both for energy and derivatives, displays a memory requirement that grows cubically with the system size. This is related to the storage of the B_{ia}^{P}, Y_{ia}^{P} and Γ_{ia}^{P} quantities, while all the other intermediates require at most an $O(N^2)$ memory. An important feature of a parallel algorithm is that, not only the computation, but also the required storage space per task is reduced by increasing the number of processes.

Reported in Table 6.1 is the amount of memory that needs to be allocated per MPI task for the storage of the most relevant intermediates. All the cubically demanding quantities are distributed over the total amount of process (N_p) or, for β_{ia}^{P} and X_{ia}^{P}, within the large number of members of each replication group. The algorithm is designed such that the computation, at the process level, involves only the allocation of quadratic intermediates. Nevertheless, these $O(N^2)$ quantities can anyways require a relatively large amount of memory, for example, the generation of the $(ia|jb)_{\mathrm{RI}}$ integrals implies the product of $v \times N_a$ matrices that, even for medium size systems, would need hundreds of Mb. That's the reason why the group has been introduced, in order to share these objects

over more processes. In fact, as shown in Table 6.1, all the quadratically demanding quantities (except for the small $P_{ij}^{(2)}$ matrix) require an amount of memory that is reduced increasing the group size N_w. Moreover, since the actual implementation is based on a hybrid OpenMP/MPI scheme, a similar gain can be achieved by increasing the number of threads employed per MPI task. This leads to more memory per MPI task without significantly loss of computational efficiency. Which of the two strategies is to be prefered is not obvious since it depends on many aspects such as the machine architecture and the implementation of the parallel libraries. As a role of thumb, using more processes per group in general leads to better workload distribution within the group, while more threads per MPI task gives better memory management.

The computation of the exchange contribution for each CPHF-like update require the contraction with 4-index ERI's over atomic orbitals $(\mu\nu|\lambda\sigma)$. These are usually calculated only at the first cycle of the procedure and then reused for the subsequent steps. For dense systems employing large basis set, that is, situations for which the integral screening is not very effective, the storage of these integrals can exceed the amount of memory available per process. In these cases only the largest possible number of ERI's are stored, while the remaining part is computed on the fly at each iteration.

In summary, the described algorithm is a scalable implementation of the RI-MP2 energy derivatives displaying, with respect of the system size, cubic, quartic and quintic requirements respectively for the memory, communication and computation.

6.4 Benchmark Calculations

6.4.1 Computational Details

Basis Sets, Thresholds and Pseudopotentials

The RI-GPW methods as implemented in CP2K[90] have been employed for all calculations in this manuscript. The correlation energy calculations are based on pseudopotentials of the form suggested by Goedecker, Teter and Hutter (GTH) in Ref. [101] but specifically parameterized for the methods employed to converge the wavefunction at the SCF level (HF or DFT). The same primary and auxiliary basis used in our previous works has been adopted [44, 131]. These are labeled as cc-DZVP, cc-TZVP and cc-QZVP, denoting double, triple and quadruple zeta quality respectively. They consist in valence-only correlation consistent type [102, 103] basis sets, generated for being used with these pseudopotentials. The Hartee-Fock exchange calculations have been performed employing our robust Γ-point implementation [104, 105] that allows for stable calculations in the condensed phase [105, 106]. The Schwarz screening threshold for the HF calculations is in the range $10^{-8} - 10^{-10}$ for the energy, while for the related derivatives the threshold is in general relaxed of one order of magnitude. Periodic calculations require a truncated Coulomb operator[105], using approximately half the length of the smallest edge of the simulation cell as truncation radius. The threshold for the SCF convergence was 10^{-6} or tighter. The PW cutoff for the HF and DFT (Perdew-Burke-Ernzerhof, PBE [170]) part of the calculations was $E_{cut} = 800$ and $E_{cut} = 1200$ Ry respectively to guarantee convergence of the exchange-correlation term. The correlation energy calculations employed high quality PW cutoffs of $E_{cut} = 300$ Ry, $E_{cut}^{rel} = 50$ Ry, $\epsilon_{filter} = 10^{-12}$, and $\epsilon_{grid} = 10^{-8}$, unless

mentioned otherwise. The convergence threshold for the Z-vector equations, measured as the norm of the residual vector, was 10^{-5} or tighter. Cluster boundary conditions have been adopted for solving the Poisson equation [95] in the case of gas phase systems. Pseudopotentials as well as all basis sets employed here, can be retrieved from Ref. [131].

For all the considered crystals, supercells have been generated by replicating the unit cell, so that the smallest edge was larger than 9Å, in order for the Γ-point approximation to be reasonable. The experimental geometries of the molecular crystals have been retrieved from the Cambridge Structural Database (CSD)[108], the structural data, together with the supercell used in the calculation and the CSD refcode, can be recovered from our previous works [44, 131] (see also Figure 3.4). For both geometry and cell optimizations no symmetry constrains have been considered, and the latter have been performed under ambient pressure. The convergence thresholds have been set to $3.0E-3$ and $1.5E-3$ Bohr respectively for the maximum and root mean square (RMS) of the geometry changes, $4.5E-4$ and $3.0E-4$ Hartree·Bohr^{-1} respectively for the maximum and RMS of the forces, while a pressure tolerance of 100 bar has been considered for the cell optimization.

The counterpoise (CP) corrected cohesive energy per molecule at a given volume V and for a given basis has been computed as [107, 111]

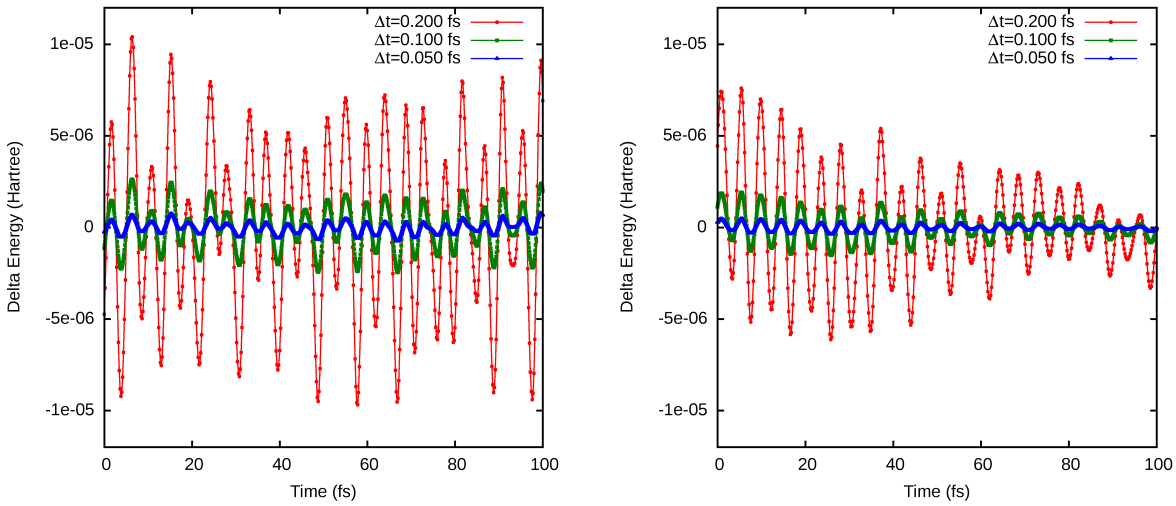
$$E_{coh}^{CP}(V) = \frac{E_{supercell}(V)}{N_{mol}} - E_{mol}^{gas} - E_{mol+ghost}^{crystal}(V) + E_{mol}^{crystal}(V). \tag{6.52}$$

Here, N_{mol} is the number of molecules per supercell, $E_{supercell}(V)$ the total energy of the supercell, and $E_{mol+ghost}^{crystal}(V)$, $E_{mol}^{crystal}(V)$, and E_{mol}^{gas} the total energy of an isolated molecule in either the crystal geometry ($E_{mol+ghost}^{crystal}(V)$ and $E_{mol}^{crystal}(V)$) or a gas phase geometry (E_{mol}^{gas}). $E_{mol+ghost}^{crystal}(V)$, includes ghost atoms from the 12 nearest neighbor molecules for NH_3 and CO_2, and from the first coordination shell in all other cases. The gas phase geometries have been obtained by relaxation at the RI-MP2 level.

6.4.2 Validation

In order to validate that forces and stress are computed consistently to the way the RI-MP2 energy is calculated, a series of short Born-Oppenheimer molecular dynamics (BOMD) simulations have been run with different time step Δt, employing the Velocity-Verlet algorithm for the integration of the equations of motion. The simulations have been performed in the microcanonical ensemble (NVE), Figure 6.5a, and in the isobaric-isothermal ensemble (NpT), Figure 6.5b. In the former only the forces acting on the atoms have to be computed while in the latter also the calculation of the stress tensor is required. The model system is made of 4 NH_3 in a cubic box with periodic boundary conditions and employing the cc-DZVP basis.

Within the Velocity-Verlet integration scheme, the total energy of an equilibrated system fluctuates around the average value with a standard deviation σ_E that is expected to be proportional to the square of the time step employed in the simulation, $\sigma_E \propto \Delta t^2$, meaning that, if Δt is halved, then σ_E is reduced roughly of a factor four. This is of course holds only in the case for which the forces, from which the accelerations are obtained, are computed as exact derivatives of the potential energy. The energies obtained from the BOMD trajectories are reported in Figure 6.5, qualitatively showing that the magnitude of the fluctuations is roughly reduced of a factor four every time Δt is halved. More



(a) **(b)**

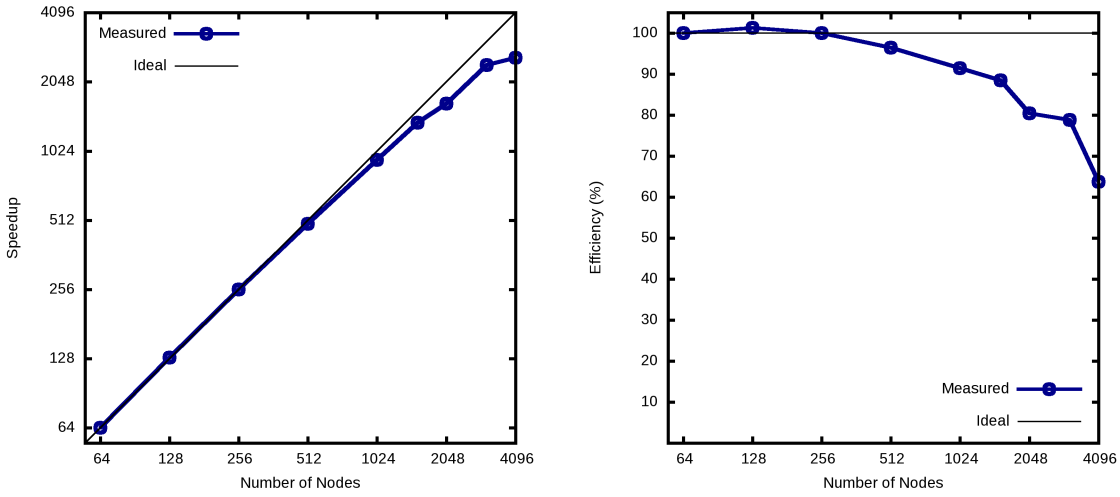
Figure 6.5: Energy fluctuation with respect to the average during a sequence of Born-Oppenheimer molecular dynamics simulations with periodic boundary conditions as a function of the time step Δt. The results are obtained employing the microcanonical ensemble (NVE) and the isobaric-isothermal ensemble (NpT) for (a) and (b) respectively. In both cases the system is made of 4 molecules of NH_3 in a cubic box with the cc-DZVP basis.

precisely the value of σ_E calculated for the NVE and NpT runs are $0.31, 1.2, 4.7$ and $0.18, 0.72, 2.9$ μHartree, respectively for time step of 0.2, 0.1 and 0.05 fs. These results are thus confirming the correctness of the RI-MP2 energy derivative implementation. We find that this approach is a stronger check than the mere comparison with numerical derivatives. For example, a large set of configurations are sampled making possible to track the propagation of possible small error that may not be detected by numerical differentiation.

6.4.3 Performance of the Methods

The parallel performance of the algorithm for calculating the RI-MP2 energy gradients and stress has been measured for a system made of 64 bulk water molecules in a cubic box with PBC at experimental density. The cc-TZVP basis has been employed resulting in 265 occupied orbitals, 3648 primary and 8704 auxiliary basis functions. The measured time includes all operations described in the previous sections, excluding only the contraction of $P^{(2)}_{pq}$ and $W^{(2)}_{pq}$ with the skeleton derivatives of the Fock and Overlap matrix elements, that is considered as part of the calculation of the HF energy derivatives. This means that also the solution of the Z-vector equation has been traced. In this respect, due to the limited amount of memory available for the smaller run, the AO-ERI's computed at the SCF level could not be kept in core during the calculation of the RI-MP2 specific quantities, and their recomputation is thus necessary before solving the Z-vector equation.

The speed-up and efficiency measured on a CRAY-XC30 machine are reported in Figure 6.6. This machine mounts a GPU on each node, but for the actual measurements



(a) **(b)**

Figure 6.6: Speed up (a) and efficiency (b) with respect to 64 nodes for the calculation of the RI-MP2 energy gradients and stress of 64 bulk water molecules (cc-TZVP basis). Calculation performed on a CRAY-XC30 machine, each node consists of 8 processes.

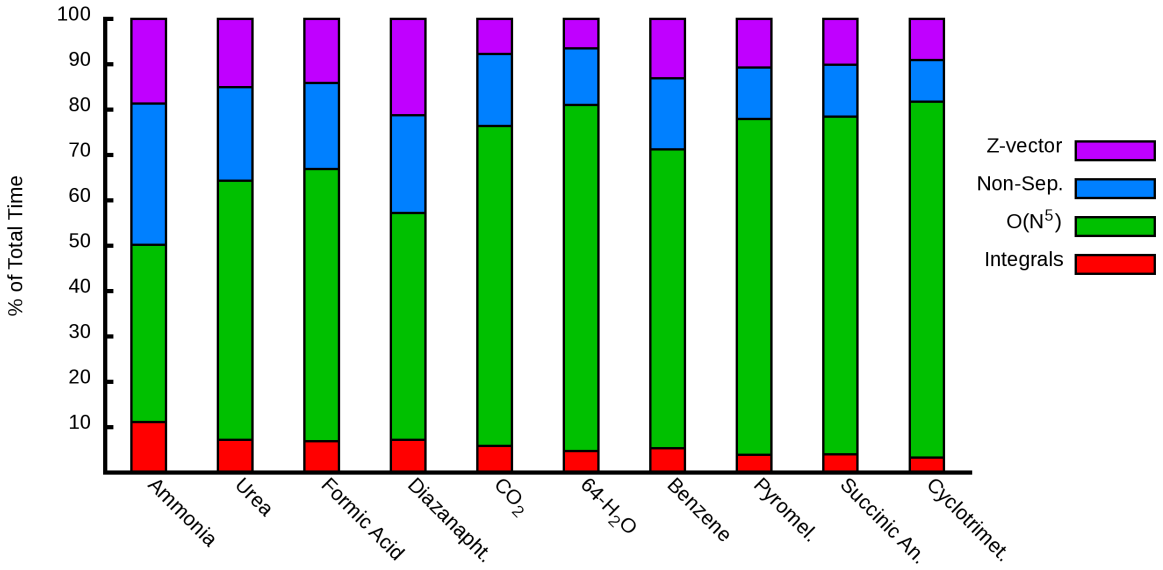


Figure 6.7: Relative time, express in term of percentage, spent in each of the most relevant part of the algorithm for the same benchmark calculations reported in Table 6.2. The meaning of the label in the legend stand for, Integrals = evaluation of $V_{PQ}^{-1/2}$ and B_{ia}^{P} (Section 6.3.1); $O(N^5)$ = evaluation of the $O(N^5)$ scaling intermediates (Section 6.3.2); Non-Sep. = evaluation of the non-separable contributions to $E_{\mathrm{RI}}^{(2)x}$ (Section 6.3.3); Z-vector = solution of the Z-vector equations, assembly of $P_{pq}^{(2)}$ and $W_{pq}^{(2)}$ (Section 6.3.4).

the usage of these devices has not been exploited. The algorithm displays good parallel scalability resulting in an efficiency higher than 80% for almost the whole range. At the

	o	n	N_a	t_{tot}	t_D	$\frac{t_D}{t_E}$	t_D^{GPU}	$\frac{t_D}{t_D^{\text{GPU}}}$
NH_3	128	2272	5312	3.15	1.53	4.20	1.47	1.04
U	192	2752	6784	5.97	3.58	4.59	2.89	1.24
FA	216	2760	6912	5.83	3.87	4.28	2.95	1.31
D	192	2992	7520	12.84	5.27	5.15	4.26	1.24
CO_2	256	2784	7296	7.94	4.99	4.15	3.50	1.43
H_2O	256	3648	8704	10.17	9.34	4.00	5.85	1.60
B	240	4128	10176	23.01	13.77	4.45	8.81	1.56
PD	312	3936	10208	28.96	17.48	4.13	9.80	1.78
SA	304	4144	10432	27.00	19.29	4.26	10.94	1.76
CT	336	4152	10560	29.71	22.30	4.16	11.97	1.86

Table 6.2: Benchmark calculations for the RI-MP2 energy gradients and stress, time in min.@CRAY-XC30, 4096 processes, 512 GPU. U = Urea, B = Benzene, FA = Formic Acid, SA = Succinic Anhydride, D = 2,3-Diazanaphthalene, PD = Pyromellitic Dianhydride, CT = Cyclotrimethylene-Trinitramine, H_2O = 64 bulk water molecules, NH_3 = Ammonia Crystal (32 molecules), CO_2 = Carbon Dioxide Crystal (32 molecules). o, n and N_a represent the number of occupied orbitals, basis functions, auxiliary basis functions respectively. The reported timings represent: t_{tot} = Total time for computing HF and RI-MP2 energy, gradients and stress; t_D = time for computing RI-MP2 energy, gradients and stress; $\frac{t_D}{t_E}$ = ration between t_D and the time for computing only the RI-MP2 energy; t_D^{GPU} = the same as t_D but employing GPU; $\frac{t_D}{t_D^{\text{GPU}}}$ = observed speed-up when using GPU.

full scale-out (32768 processes) the time for computing the RI-MP2 energy gradients and stress is 106 seconds. The relatively large drop in efficiency observed in going from 3072 to 4096 nodes is related to the scarce number of ij pairs processed by each MPI task in the latter, such that the time spent in computation becomes of the same order of the overheads related to communication.

In Table 6.2 the timing for different benchmark calculations, obtained employing 512 nodes of a CRAY-XC30 machine, are reported, in this case also the impact of the usage of the GPU's has been considered. In general, for the actual implementation, the GPU's have been used to accelerate all the steps that are performed in the algorithm as matrix multiplication. This is of particular advantage for the RI-MP2 method since the expected most computationally intense part, *i.e.* the $O(N^5)$ steps, are all accomplished in this way.

For the sake of completeness, in Table 6.2 also the time t_{tot} necessary for the evaluation of the energy gradients and stress of the total energy (HF + RI-MP2) is reported. At the Hartree-Fock level the most expensive operations are related to the update of the Fock matrix with the exact exchange contributions, that involves the calculation of the AO-ERI's and relative derivatives. For the reported benchmarks, computing the derivatives of the RI-MP2 energy results in a percentage of the total time that grows systematically increasing the system size up to 75% for the largest case. The exception to that is the 64

bulk water case, for which the Schwarz screening, in particular for the $(\mu\nu|\lambda\sigma)$ derivatives, is particularly effective, resulting in a small time spent at HF level.

The total time necessary for calculating the RI-MP2 energy gradients and stress (t_D), reported in Table 6.2, is of the order of minutes for all the cases and results to be a factor between 4 to 5 times larger (t_D/t_E) than what required for the calculation of the RI-MP2 energy only. According to the analysis done in Section 6.3.2 in the limit of very large system, *i.e.* when the $O(N^5)$ steps are by far the most time consuming part of the total computation, the ratio t_D/t_E is expected to be between 3 to 4. This is just the theoretical limit, since the calculation of the integral derivatives as well as the solution of the Z-vector equations will always give a non negligible overhead to the calculation.

The relative time spent in each part of the algorithm, for the different benchmark calculations, are reported as percentage of the total in Figure 6.7. For all cases, except the Ammonia crystal, the time spent for calculating the $O(N^5)$ intermediates is more that 50% of the total, reaching almost 80% for the largest case (CT). The computation of the RI specific quantities, $V_{PQ}^{-1/2}$ and B_{ia}^{P} is the cheapest operation requiring less than 10% of the total computational effort for all calculations. The evaluation of the non-separable contributions to $E_{\mathrm{RI}}^{(2)^x}$ is dominated by the calculation of the 3-center integrals and associated derivatives, and results to be roughly a factor 3 more expensive than the computation of $V_{PQ}^{-1/2}$ and B_{ia}^{P}. In this respect, it has to be noted that the computation of the non-separable contributions to the RI-MP2 stress tensor takes around 30% of time spent in this part and this computation is avoided in the case for which only gradients are required. The remaining part is associated with the solution of the Z-vector equations that can require a variable percentage of the overall time according the the number of iteration necessary to reach convergence. For all the reported cases, the amount of memory was enough for keeping in core the AO-ERI's computed at the SCF level during the evaluation of the RI-MP2 specific quantities. This allowed to avoid their recomputation for the solution of the Z-vector, making this operation less computationally demanding.

Finally the time for the calculation of the RI-MP2 energy gradients and stress by exploiting the GPU's for the operations performed as matrix multiplication is reported in Table 6.2 labeled as t_D^{GPU} together with the observed speedup compared to the only CPU case (t_D/t_D^{GPU}). As shown in the table, the speedup is modest for the smaller cases while approaching a factor 2 for the larger ones. Focusing on the largest benchmark calculation (CT), the observed speedup for the overall calculation is roughly 1.9, but what observed for each of the single matrix multiplication is in general greater than 4.

6.4.4 System Size Scaling

In order to verify the cost models presented in the previous sections, the time for each of the individual steps of the algorithm has been measured for increasing system sizes. The test system is based on a supercell containing 32 bulk water molecules with a cc-DZVP basis set that has been replicated in one dimension up to 5 times. The results are reported in Figure 6.8 where the obtained timings have been fitted with the function $y = bx^a$, yielding the measured scaling exponent a associated with each different step.

In Figure 6.8 the labels "Integrals" and "Non-Sep." refer to all operations described in Section 6.3.1 and 6.3.3 respectively. The evaluation of the $O(N^5)$ scaling intermediates (Section 6.3.2) have been traced in more details, reporting separately the measured timings

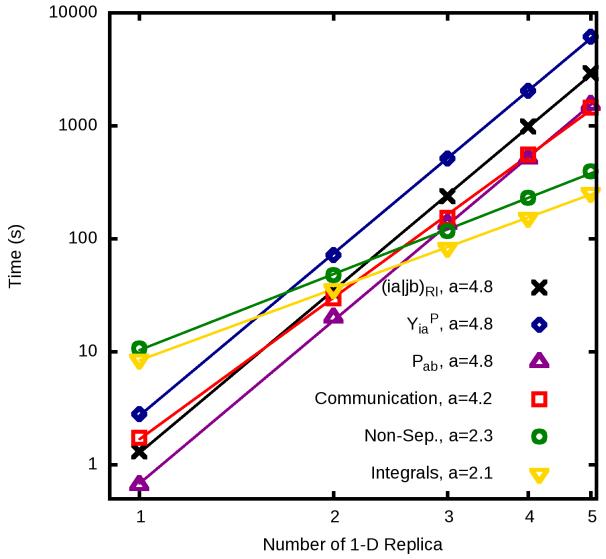


Figure 6.8: Time spent in the various significant part of the algorithm for the calculation of the RI-MP2 energy gradients, as a function of the number of replicas of the supercell, containing 32, 64, 96, 128 and 160 molecules of H_2O respectively (cc-DZVP basis). Timing measured on a CRAY-XK7 machine employing 2400 processes without GPU's. Lines represent a linear two-parameter fit of the form $y = bx^a$. The values of a for each operation are reported in the legend.

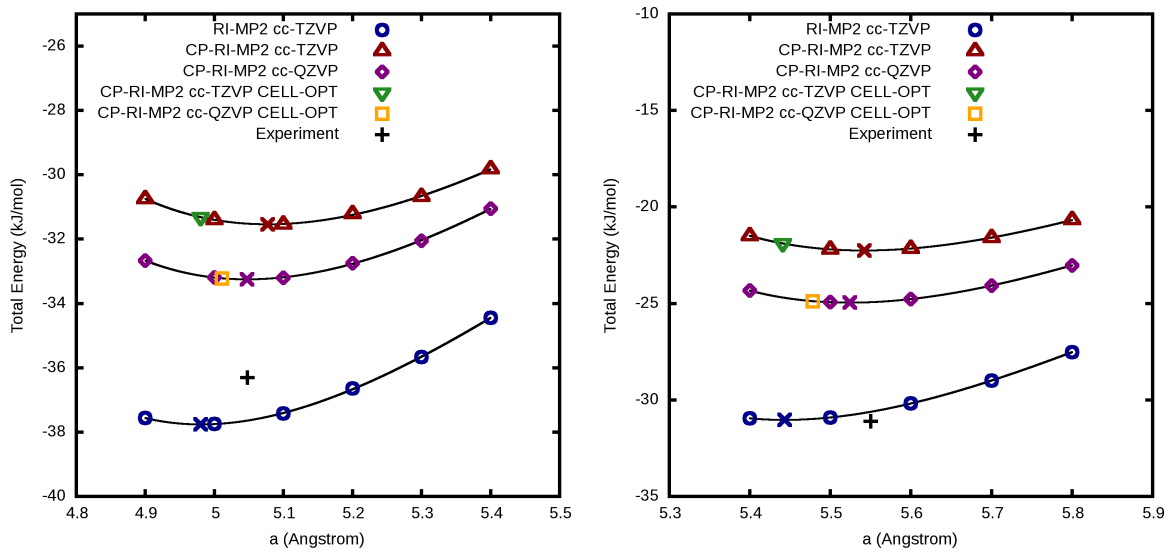
for the generation of the $(ia|jb)_{\mathrm{RI}}$ integrals, the update of Y_{ia}^P and $P_{ab}^{(2)}$, and communication. Note that, the solution of the Z-vector equations in this case, due the combination of the system topology and small basis, takes a negligible time with respect to the total and thus has not been reported in the plot.

The observed scaling for "Integrals" and "Non-Sep." is 2.1 and 2.3 respectively showing that the integration of the electrostatic potential over pairs of basis elements $\mu\nu$ is the dominant part within the tested sizes. This operation is in fact expected to scale as $O(N^2)$, while the additional parts, such as integral transformation, scaling as $O(N^4)$, make the exponent a slightly larger than 2. This effect is more pronounced for the latter compared to the former due to the higher number of $O(N^4)$ steps performed in the update of the non-separable part of $E_{\mathrm{RI}}^{(2)^x}$.

For the evaluation of the $O(N^5)$ intermediates ($(ia|jb)_{\mathrm{RI}}$, Y_{ia}^P and $P_{ab}^{(2)}$) the observed scaling is in all cases 4.8, while communication has a measured a of 4.2. This is in agreement with the performance models derived for these operations, that is 5 for the former and 4 for the latter. It is interesting to note the relative timing of each of the $O(N^5)$ steps, for which the update of Y_{ia}^P is roughly a factor 2 compared to the generation of $(ia|jb)_{\mathrm{RI}}$, while this is again a factor 2 compared to the update of $P_{ab}^{(2)}$. The reason for the first observation is related to the fact that, for a given ij pair, the update of Y_{ia}^P and the generation of $(ia|jb)_{\mathrm{RI}}$ require for both $O(v^2N_a)$ operations but for Y_{ia}^P this is performed 2 times (for i and j respectively). The update of $P_{ab}^{(2)}$ requires $O(2v^3)$ for each ij and since in this case $N_a \simeq 4v$ the observed time scales as $O(v^2N_a/2)$ that is half than what needed for generating $(ia|jb)_{\mathrm{RI}}$.

6.4.5 Applications

Solid NH_3 and CO_2



(a) **(b)**

Figure 6.9: Location of the minima for NH_3 (a) and CO_2 (b), computed at the RI-MP2 level of theory with different basis sets obtained with different approaches. The lattice parameter optimization curves have been fitted with a third order Birch-Murnagham equation, the crosses represent the location of the minimum point for each curve. CP means that the cohesive energy have been counterpoise corrected.

Ammonia and carbon dioxide molecular crystals represent two simple benchmark systems useful for judging the performance of a method. The dominant interactions in the two cases are very different in nature, being weak hydrogen bond for NH_3 and purely van der Walls for CO_2. These systems have been extensively investigated both experimentally than theoretically. Concerning the theoretical studies, many of them are MP2 theory based methods such as periodic-canonical MP2 [44], periodic-local MP2 (LMP2) [107], incrementally corrected LMP2 [246], embedded many-body expansion [247, 248] and hybrid QM/MM fragment method [249].

Two approaches can be used for calculating the equilibrium lattice parameter (a) and cohesive energy (E_{coh}) of these crystals. The first approach (direct method) is to perform a cell optimization followed by the calculation of E_{coh} for the equilibrium structure. The second one (indirect method) consists in the optimization of the geometry at various fixed volumes from which the equilibrium quantity a and E_{coh} are derived by fitting employing, *e.g.* a third order Birch-Murnagham equation. The former is computationally more efficient since a single optimization has to be carried out, moreover it allows to gain more information on the local molecular structure at equilibrium. On the other hand, this approach suffers from basis set superposition error (BSSE) that can be particularly large within MP2 theory. This drawback can be remedied by increasing the basis set, or, in the indirect method, by considering counterpoise corrected energies. Both approaches have

	NH_3				CO_2		
	a	$r_{\text{N-H}}$	$\angle_{\text{H-N-H}}$	E_{coh}	a	$r_{\text{C-O}}$	E_{coh}
Opt-CP (PBE-D3)	5.00	1.027	107.5	-44.97	5.73	1.171	-26.50
Fit-Opt	4.98	-	-	-37.76	5.44	-	-31.04
Fit-Opt-CP	5.08	-	-	-31.54	5.54	-	-22.26
Fit-Opt-CP (QZ)	5.05	-	-	-33.25	5.52	-	-24.96
Opt-CP	4.98	1.017	107.4	-31.32	5.44	1.166	-21.90
Opt-CP (QZ)	5.01	1.017	107.2	-33.23	5.48	1.168	-24.88
Exp.	5.048	1.01 - 1.06	107.5	-36.3	5.55 - 5.62	1.155 - 1.12	-31.1

Table 6.3: Equilibrium cohesive energy (E_{coh} in kJ/mol per molecule) and structural properties (lattice parameter a and bond length in Å, angles in degree) for the NH_3 and CO_2 crystals calculated employing different methods. Except when specified otherwise, the basis set is cc-TZVP and the level of theory RI-MP2 (QZ stands for cc-QZVP basis). CP means that E_{coh} is counterpoise corrected, Opt denotes full cell optimization while Fit-Opt refers to the results obtained by fitting the curves shown in Figure 6.9 (third order Birch-Murnagham equation). Experimental values from Ref. [107] (see also Ref. [121, 243, 244]) and Ref. [122, 123, 245] and references therein for NH_3 and CO_2 respectively.

been considered employing the cc-TZVP and cc-QZVP basis, the computed equilibrium properties are summarized in Table 6.3.

As shown in Figure 6.9, at the triple zeta level, the lattice parameter optimization curves are calculated both with and without CP correction. The position of the obtained minima, indicated by the crosses in the plots, shows clearly the large discrepancies between the two approaches. As expected, for both systems, without CP correction the crystals result overbind with shorter equilibrium lattice parameter and larger cohesive energy. The CP correction fixes this issue giving values for a closer to the experimental one but E_{coh} in general higher.

The cell relaxation provides converged structures that preserve the cubic symmetry of the crystals within the numerical accuracy of the method. With the cc-TZVP basis, the obtained lattice parameters from cell optimization (Opt) are essentially the same as those obtained from the non-CP curve optimizations (Fit-Opt) and thus substantially shorter than those evaluated with the CP correction (Fit-Opt-CP). This divergence is effectively reduce when using the cc-QZVP basis, showing a clear trend in the convergence. In fact, by inspection of Figure 6.9, it can be noted that the Fit-Opt-CP approach tend to converge, with respect to the basis set, from larger values of a opposite to the case of the cell optimization.

At the quadruple zeta level the lattice parameters obtained from the cell optimization are 5.01 and 5.48 Å respectively for NH_3 and CO_2, with associated cohesive energies of -33.2 and -24.9 kJ/mol. With the same basis the CP curve optimization procedure, Fit-Opt-CP(QZ), gives similar results for E_{coh}, but slightly larger values of a being respectively 5.05 and 5.52 Å. From the observations previously stated it can be concluded that, for the supercell considered here, the complete basis set limit for the equilibrium lattice parameter should be within these values, that is between 5.01-5.05 and 5.48-5.52 Å, respectively for NH_3 and CO_2.

For ammonia, good agreement is found with the values reported by Maschio *et al.*

[107] obtained with the aug(d,f)-TZPP basis, that is $a = 5.02$ Å and $E_{coh} = -36.6$ kJ/mol, while a larger deviation in the lattice parameter is observed for carbon dioxide ($a = 5.59$ Å and $E_{coh} = -26.6$ kJ/mol). A better agreement for the lattice constant a of the CO_2 crystal is obtained when comparing with the value of 5.52 and 5.46 Å reported respectively by Bygrave *et al.* [247] and Sode *et al.* [248], calculated with a CP augmented triple-zeta basis for the former and augmented quadruple-zeta basis for the latter. In this case also the optimized C-O bond length match the values reported by these authors.

As a comparison, also the results obtained with the PBE functional including the Grimme D3 [124] correction have been reported. For ammonia good agreement is found between RI-MP2 and PBE-D3 in the lattice constant, while the cohesive energy results much larger for the latter. On the other hand, for CO_2 the E_{coh} is estimated roughly the same but the value of a is around 5% larger than that calculated with RI-MP2.

Caution has to be exercised when comparing the obtained results with experiments since many effects, such as temperature dependence and zero-point vibrational energies, are neglected. For both crystals, at the QZ level, the lattice parameters are less than 2% shorter than the experimental ones while the cohesive energies display larger deviations.

Molecular Crystals

Geometry and cell optimization at the RI-MP2 level employing the cc-TZVP basis have been carried out for a set of molecular crystals. This set includes the crystal of Urea (U), Formic Acid (FA), Benzene (B), Pyromellitic Dianhydride (PD), Succinic Anhydride (SA) and Cyclotrimethylene-Trinitramine (CT). The dominant intermolecular interactions for these crystals cover a large range, from hydrogen-bond to dipole-dipole till purely van der Waals. Also the size of the molecules across the investigated set are quite different going from a minimum of 5 (FA) up to 21 atoms for the largest case (CT). For the relaxed structures the counterpoise corrected cohesive energy has been computed, the obtained results, compared with the experimental values, are reported in Table 6.4.

For all cases, the cell optimization preserves the orthorombic symmetry of the crystals, roughly keeping the experimental aspect ratio. The obtained lattice parameter are in all cases underestimated compared to experiment, with a deviations ranging between 1 up to 7%. This reflects into much larger discrepancies for the cell volumes for which a maximum deviation of 15% is observed for the Benzene crystal. A slightly smaller error, around 8% for the volume, is observed for the Urea and Formic Acid crystals, cases for which the intermolecular interactions are mainly of dipole-dipole and hydrogen-bond types. This difference can be rationalized by inspection of the S22 set [130], for which it is shown that MP2 is in general giving poor performance for the complexes with predominant dispersion contribution, such as the case of Benzene, while the the results are usually better for hydrogen bonded complexes [130, 171].

As in the cases of solid NH_3 and CO_2, the cell optimization at the triple-zeta level is clearly affected by the BSSE. In order to estimate to which extent the observed error in the converged lattice parameters is due to BSSE or to intrinsic limitation of the MP2 theory, a full cell relaxation employing the cc-QZVP has been performed for the Formic Acid crystal. The obtained lattice constants at the quadruple-zeta level are $a = 10.20$, $b = 3.41$ and $c = 5.33$ Angstrom while the computed cohesive energy is -59.3 kJ/mol. Increasing the basis set is thus reducing the error in the equilibrium volume to 4.2%. In

	Geo-Opt	Cell-Opt			Exp.		
	E_{coh}	abc	V	E_{coh}	abc	V	E_{coh}
		5.45			5.65		
U	-97.1	5.45	138	-96.6	5.65	150	-92
		4.64			4.70		
		10.06			10.24		
FA	-55.6	3.36	179	-54.9	3.54	194	-68
		5.31			5.36		
		7.14			7.40		
B	-58.8	8.78	400	-63.0	9.44	473	-45
		6.39			6.78		
		10.16			10.79		
PD	-125.7	10.18	754	-130.4	10.79	863	-83
		7.29			7.41		
		5.14			5.43		
SA	-82.6	6.57	382	-84.8	6.97	443	-81
		11.31			11.72		
		12.88			13.18		
CT	-116.6	11.05	1452	-115.8	11.57	1634	-112
		10.21			10.71		

Table 6.4: Counterpoise corrected cohesive energy E_{coh} (kJ/mol) equilibrium volume V (Å^3) and lattice parameters abc (Å) for different molecular crystals calculated after structural relaxation at the RI-MP2 level of theory employing the cc-TZVP basis. The meaning of the labels are: U = Urea, FA = Formic Acid, B = Benzene, PD = Pyromellitic Dianhydride, SA = Succinic Anhydride, CT = Cyclotrimethylene-Trinitramine. Geo-Opt refers to geometry optimization at experimental volume while Cell-Opt stands for full cell relaxation. The experimental E_{coh} are obtained from sublimation enthalpies $\Delta H(s)$ with opposite sign. For the experimental date see Ref [44, 111], as well as http://webbook.nist.gov/chemistry/ and Cambridge Structural Database [108].

particular, while for the a and c vectors the agreement is fairly good, a large deviation is observed for the b lattice parameter, such that the discrepancy in the volume is almost completely determined by this underestimation. The solid formic acid consists of infinite chains of molecules linked by hydrogen bonds, while interchain interactions are dominated by dispersion and weak C-H$\cdots$O contacts. The weak intermolecular interactions act along the cell vectors a and b, while the hydrogen bonded formic acid molecules form infinite chains approximately oriented along cell vector c [250, 251]. The observed compression observed along the b vector can be interpreted as the result of the overbinding of dispersion interactions at the MP2 level. Note that the cell optimization for FA at the PBE level with the cc-TZVP basis gives $a = 11.19$, $b = 4.20$ and $c = 5.24$ Å, in good agreement with

$a = 10.91$, $b = 4.11$ and $c = 5.28$ Å, reported by Tosoni *et al.* [250] obtained employing the Ahlrichs' TZP basis.

The CP corrected cohesive energies E_{coh} reported in Table 6.4 have been computed for the relaxed structures obtained after geometry optimization at experimental volume and cell optimization. By comparison of the calculated values appears that, even if a strong structural relaxation take place, this doesn't reflect into a large variation of E_{coh} which results to be only slightly different for the two cases. This is a direct consequence of the weak binding interactions that dominate in these systems, giving rise to relatively flat potential energy surfaces with respect to the cell parameters. Moreover, due to the shrinking of the volume during the cell relaxation, the structure of the isolated molecules are usually less stable when extracted from the geometries obtained from Cell-Opt than those obtained from Geo-Opt. This effect partially compensates the gain in energy due the optimization of the cell parameters.

Even if care has to be considered when comparing E_{coh} with experimental sublimation enthalpies $\Delta H(s)$, it can be noted that at the MP2 level good agreement is found when the crystal are bounded with mixed electrostatic-dispersion interactions such in the case of Urea, Succinic Anhydride and Cyclotrimethylene-Trinitramine [252]. On the other, for crystals such as Benzene and Pyromellitic Dianhydride, purely bounded with van der Waals dispersion interactions, large deviations are observed, resulting in general in an overestimation of the cohesive energy.

6.5 Conclusions

In this chapter, the equations for calculating the derivatives of the MP2 energy in the framework of the Resolution of Identity Gaussian and Plane Waves method have been derived and presented in details. The central aspect in the computation of the derivatives of the correlation energy within the RI-GPW approach is the dual representation of the RI fitting density in term of Gaussian and Plane Waves auxiliary functions. The latter representation is equivalent to expressing the electrostatic densities over regular grids in space. This allow the straightforward conversion of these quantities into the associated potentials by solving the Poisson equation in G space and exploiting Fourier transformations for switching between direct and reciprocal representations. In this way the evaluation of the integral derivatives is accomplished consistently to the way the energy is calculated. This approach is of general validity and it has been applied to the calculation of the forces acting on the atoms (gradients) as well as for the derivative with respect to the cell volume (stress tensor).

For the presented scheme a massively parallel algorithm has been designed displaying, with respect of the system size, cubic, quartic and quintic requirements respectively for the memory, communication and computation. All these requirements scale increasing the number of processes. The implementation is based on hybrid OpenMP/MPI scheme for which the parallelization is achieved by distributing the work over subgroup of processes rather than over single task. This allowed to achieved a more flexible memory management and reduced communication without loss of computational efficiency. The measured performance displays excellent parallel scalability and efficiency up to thousands of nodes. Moreover, in the actual implementation the computationally more demanding

part, that is the $O(N^5)$ steps, are accelerated by employing GPU's showing a gain of almost a factor two compared to the standard only CPU case for large systems.

Several benchmark calculations have been reported both for validating the theoretical and methodological aspects. It has been shown that the presented scheme is efficient, accurate and robust especially for systems in the condensed phase. The effort for the calculation of the derivatives at the RI-MP2 level is between 4 to 5 times more expensive than computing only the energy. Geometry optimization as well as full cell relaxation have been performed for a large variety of molecular crystals. The obtained results are in general good agreement with both previously reported calculations and experimental data. Furthermore, the actual implementation allows for fully exploit the computational power of new generation supercomputers, such that the derivatives of the RI-MP2 energy can be performed within minutes for systems containing hundreds of atoms and thousands of basis functions.

The methodology presented here poses the basic machinery that can be extended for the calculation of the energy derivatives evaluated at the double-hybrid density functional [253–255] and random-phase approximation [186] level, for which (N^4) scaling implementations have been reported.

Chapter 7

Bulk Liquid Water at Ambient Temperature and Pressure from MP2 Theory [177, 256]

MP2 provides a good description of hydrogen bonding in water clusters and includes long range dispersion interactions without the need to introduce empirical elements in the description of the interatomic potential. To assess its performance for bulk liquid water under ambient conditions, an isobaric-isothermal (NpT) Monte Carlo simulation at the second-order Møller-Plesset perturbation theory level (MP2) has been performed. The obtained value of the water density is excellent (1.02 g/mL) and the calculated radial distribution functions are in fair agreement with experimental data. The MP2 results are compared to a few density functional approximations, including semi-local functionals, hybrid functionals, and functionals including empirical dispersion corrections. These results demonstrate the feasibility of directly sampling the potential energy surface of condensed phase systems using correlated wavefunction theory, and their quality paves the way for further applications.

7.1 Introduction

Understanding the structural and electronic properties of liquid water at ambient conditions is a major challenge in condensed matter simulations. Water is a crucial ingredient for a large variety of systems of prime importance in basic chemistry, biology, and physics, as well as in the applied fields of catalysis and energy production. The water molecule has a large dipole moment and polarizability, is a multiple hydrogen donor and acceptor and can easily build network structures. The total cohesive energy in the condensed phase is, as a consequence of these properties, a sum of many weak interactions. Theoretical models face therefore the challenge to describe many different effects and their subtle interplay at a high precision. The development of sophisticated empirical potentials for water [257–266], allowed to gain insights into water's behavior and its thermodynamic properties [193, 267, 268], such as, density maxima, heat capacity and effects of supercooling. However, empirical models lack transferability and might fail if used under conditions away from their fitting range. Most importantly, as soon as

water takes an active role in a chemical process, either as a strongly interacting solvent, or for example as a source of protons, the electronic properties of the water molecule need to be taken into account. In this respect, first-principles methods offer the possibility to describe all the underlying physics on the same footing, simplifying the treatment of intra- and inter-molecular interactions. The capability to reproduce properties of complex systems such as liquid water can therefore be used to judge the sophistication and predictive power of a given quantum mechanical model. Density functional theory (DFT) is the most used quantum mechanical method employed for studying physical and chemical properties of condensed phase systems. Many DFT based simulation of bulk water have been reported in the literature, and in this context three main methods of sampling the phase space can be recognized [269]: the Car-Parrinello molecular dynamics (CPMD) and related variants [195, 202, 270–274], Born-Oppenheimer molecular dynamics (BOMD) [176, 203, 275–277] and Monte Carlo (MC) sampling [175].

Most of the CPMD and BOMD simulations were carried out in either the micro-canonical (NVE) or canonical (NVT) ensemble by constraining the volume to reproduce the experimental density ρ. First principles MD simulations in the isobaric-isothermal (NpT) ensemble are much less common [176, 278], in part due to the need for different integration schemes, the increased requirements on plane waves basis sets in variable cell simulations, and the long simulation time that is needed to equilibrate and to sample volume fluctuations. Moreover, in the MD sampling of the NpT ensemble, the calculation of the virial and thus stresses is required to change the volume as a response to the imbalance between internal stress and external pressure [279]. In this respect, the appealing feature of the MC method is that thermodynamic constraints are explicitly included into the acceptance rule for each trial move, solely based on the energy. This allows to perform simulations in different ensembles, *e.g.* NpT, in a relatively straightforward manner. On the other hand, an efficient sampling of phase space in MC requires smart and system dependent trial moves [280–284], making the application of the method more intricate than molecular dynamics where configurational sampling follows a general principle. Within the framework of MC, McGrath and coworkers [175] reported the first results from first-principles simulations of liquid water in the isobaric-isothermal ensemble at ambient pressure. These NpT-MC simulations performed at ambient conditions using the Becke-Lee-Yang-Parr (BLYP) [109, 110] functional gave a significantly less dense (about 20%) liquid than observed experimentally. These results were confirmed and extended by Schmidt *et al.* [176] using very similar computational setups, but within the framework of NpT-MD. In the latter work, in addition to BLYP, also the Perdew-Burke-Ernzerhof (PBE) [170] functional has been tested, and, in both cases, the influence of an empirical dispersion correction (DFT-D) [285] has been investigated. It was shown that pure PBE and BLYP are indeed underestimating the water density, but the inclusion of the dispersion correction significantly improves the results giving a density as well as oxygen-oxygen radial distribution function (RDF) closer to the experimental data. From these calculations the important role played by the van der Waals interactions in bulk water clearly emerged, and it was shown that this missing interaction in standard local functionals can be efficiently included using empirical corrections [124, 125, 285]. Furthermore, it became clear that NpT simulations are essential to quantify the quality of the intermolecular interaction potential and that a fixed simulation volume should be considered an influential constraint. Despite this progress, a truly first principles

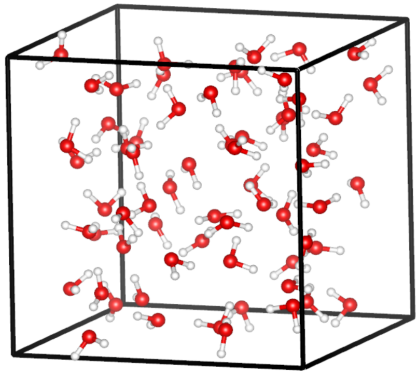


Figure 7.1: Initial configuration used for all MC simulations, it consists of 64 water molecules in a cubic box with initial cell edges of $L = 12.42\,\text{Å}$.

simulation of liquid water in the NpT ensemble is still missing.

Here, we present the results of MP2 based NpT-MC simulation of liquid water at ambient conditions. The second-order Møller-Plesset energy [13, 14] is an effective correction to the Hartree–Fock (HF) ground state energy that accounts for electron correlation effects. It recovers a relatively large part of the dynamic correlation, while the HF ground state ensures the inclusion of exact non-local exchange. Most notably, MP2 introduces dispersion interactions in a completely non-empirical way. For water clusters, MP2 produces results that are in excellent agreement with coupled cluster theory [286–289], and is one of the best methods for the WATER27 benchmark [290]. However, the advantages of MP2 come at a computational cost that is high compared to traditional DFT using local functionals. Furthermore, MP2 calculations need larger basis sets than DFT to reach a similar state of convergence. Therefore, only few applications of MP2 to condensed phase systems have been reported (see e.g. [51, 291, 292]) and condensed phase sampling has not yet been performed. The extensive calculations reported here have become possible by combining large computer resources with an efficient algorithm for the MP2 energy calculation, which we named the Resolution of Identity Gaussian and Plane Wave (RI-GPW) approach [44, 131] and implemented in the CP2K [90] program. In addition to the new MP2 results, we report also results of NpT-MC simulations obtained from a selection of density functional approximations, namely, BLYP, PBE and PBE0 [173], including empirical dispersion corrections of the D3 type [124], which have been generated to validate the approach.

7.2 Computational Details

All calculations presented have been performed with the CP2K program [90]. Hartree–Fock and DFT energy calculations employ the Gaussian and Plane Wave (GPW) method [86, 91] that makes use of a Gaussian basis to expand molecular orbitals and an auxiliary plane wave basis for the expansion of the electronic density. This dual representation allows for evaluating the Hartree contribution to the Kohn–Sham (KS) matrix in linear scaling time at full accuracy [91]. In order to have an efficient expansion of the density in plane waves, core electrons are replaced by pseudopotentials. We use dual-space pseudopotentials of the Goedecker-Teter-Hutter (GTH) type [101] specifically parameterized for the

various methods used (HF, BLYP, PBE, PBE0). The non-local HF exchange calculations have been performed employing a robust Γ-point implementation [104, 105]. To avoid divergences in the non-local HF exchange energy, the Coulomb operator is truncated at half the cell length. This allows for stable calculations in periodic boundary conditions without loss of accuracy if a truncation radius larger than $5-6$ Å is used for wide bandgap systems [105, 106]. In the case of the hybrid functional (PBE0), the calculations have been performed using the auxiliary density matrix method (ADMM) [293] that allows for an approximate calculation of the non-local exchange energy at much reduced cost. This is achieved by introducing a small auxiliary Gaussian basis that is used to expand the density matrix employed in HF exchange energy calculation. The exchange energy is then corrected for the difference between the exact and approximated density using a local exchange functional. The calculation of the MP2 energy within the RI-GPW [131] approach (RI-MP2) is closely related to the original GPW method. Here, the dual representation of the electronic density is applied to the fitting density arising from the resolution of identity approximation [69, 144, 145], where the RI fitting densities are obtained by introducing an auxiliary Gaussian basis and employing the Coulomb metric [146].

The model system consists of 64 water molecules in a cubic simulation cell under periodic boundary conditions (PBC), Figure 7.1. Initial configurations were taken from a system extensively equilibrated using the BLYP-D3 functional at $T = 295$K and experimental density $\rho = 1.0\,\mathrm{g/cm^3}$, i.e. constant volume with cell edges of $L = 12.42$ Å. All further simulations have been obtained with thermodynamic constraints set to ambient conditions, that is, $T = 295$K and $p = 1$bar. The employed basis set for all calculations is of the cc-TZVP form [44, 131]. That is a valence triple-zeta correlation-consistent type basis [102, 103], but specifically generated for pseudopotentials. The cc-TZVP basis consists of sets of $(5s5p2d1f)$ and $(5s2p1d)$ primitives contracted to $(3s3p2d1f)$ and $(3s2p1d)$ functions for oxygen and hydrogen, respectively. The inclusion of the f and d functions is necessary in order to obtain accurate results at the MP2 level. At the DFT/HF level this basis is of similar quality as the QZV3P basis, that has been shown to provide very well converged DFT results [203, 294]. The associated RI auxiliary basis (RI-cc-TZVP) [131] has been generated following the procedure proposed by Weigend *et al.* [69]. It is about 2.5 times larger than the primary basis and contains up to g-type and f-type functions for O and H, respectively (see Appendix B for more details). In the PBE0 case an additional auxiliary basis is needed for calculating the HF exchange energy with ADMM. This basis (pFIT3) contains only non-contracted Gaussian functions with primitive patterns $(3s3p1d)$ for O and $(3s1p)$ for H. The number of grid points used in the GPW scheme for the representation of the electronic density is kept constant during volume changes. The original grids are constructed for the reference cell with density $1.0\,\mathrm{g/cm^3}$ ($L = 12.42$ Å). The charge-density cutoffs (E_{cut}) for the HF and DFT calculations are 500 and 800 Ry, respectively, while for the RI-MP2 calculation a cutoff of 300 Ry is used. The higher cutoff for DFT calculations is necessary as the exchange-correlation functional is integrated on the same grid [44, 91]. We have verified for the BLYP functional that increasing E_{cut} from 800 Ry to 1200 Ry does not affect final results. With this setup, the model consists of 64 molecules (192 atoms), 256 electrons in 128 occupied orbitals, 57 primary basis functions per molecule (3648 basis functions in total), 136 auxiliary basis functions per molecule (8704 auxiliary basis functions in total), and plane wave grids of sizes 216^3, 180^3, and 135^3. Representative run times for a single energy calculation on a

Cray XE6 (32 cores per node) computer are 218 seconds for RI-MP2 on 400 nodes, 34 seconds for ADMM-PBE0 on 24 nodes, and 17 seconds for local density functionals on 16 nodes.

For all the theoretical models presented (DFT or RI-MP2) the MC settings are kept the same. The considered trial moves are: (1) atom translations, (2) molecular translations, (3) molecular rotations and (4) volume changes. The MC efficiency is improved with the presampling of moves [280, 282]. In this method, an additional inexpensive (approximated) potential is introduced and used to generate a short sequence of MC moves. The entire sequence is then accepted or rejected completely, based on the difference between the exact and approximated potential functions. In the present study, the presampling consists of 30 steps with 8 simultaneous moves of type (1), (2), or (3), selected randomly with equal probability. Volume moves are not considered for the presampling, but instead directly based on the exact potential. The ratio of volume move to presampling sequences is set to 1:9. After extensive testing, the maximum displacements for moves (1) to (3) was adjusted to give acceptance probabilities between $15 - 20\%$ for the inner moves of the presampling sequence. This choice allows for large enough configuration changes while still keeping a good acceptance rate for the exact potential. On average, 38% of the atoms in the unit cell are moved by this presampling sequence. The maximum displacement of the volume moves is tuned in order to give a 50% acceptance rate. For the DFT simulations, a classical but refitted force field based on Ref. [295] is used as approximate potential. For the MP2 simulations, a higher quality presampling based on a semi-local density functional method is employed. In particular, the PBE1W [296] functional is used with a double-zeta plus polarization basis set and including the D3 [124] empirical dispersion correction. Finally, the length of each MC simulation is reported in terms of Monte Carlo cycles. Each cycle refers to a single exact energy calculation, which thus involves either a volume move or an entire sequence of presampling moves, that can be accepted or rejected.

Psudopotentials, basis set parameters and initial geometry can be retrieved from supporting information of ref. [177].

7.3 Results

A central result of this chapter is the computed value of the density of liquid water. This quantity is obtained from averaging the fluctuating instantaneous density as obtained during the MC simulation. The instantaneous density and the corresponding running average are depicted in Figure 7.2, while the calculated average value with the associated root mean square deviations (RMSD) are reported in Table 7.1. Our MP2 result for the density of liquid water at ambient conditions is ≈ 1.020 g/mL. We consider this to be in excellent agreement to the experimental value, since it implies an error in the lattice parameters of less than 1%. We attribute this to the known quality of MP2 for describing water hydrogen bonds, and the presence of dispersion interactions. Moreover, it is known from classical and DFT based simulations that, for flexible water models, the inclusion of nuclear quantum effects leads to less structured liquid and improve the density behavior [298–301]. By observing that these effects are less pronounced in D_2O than in H_2O and that the former has a molar density 1.3% higher than the latter, we expect the inclusion of this correction to reduce further the density [300, 302]. This result stands also in sharp

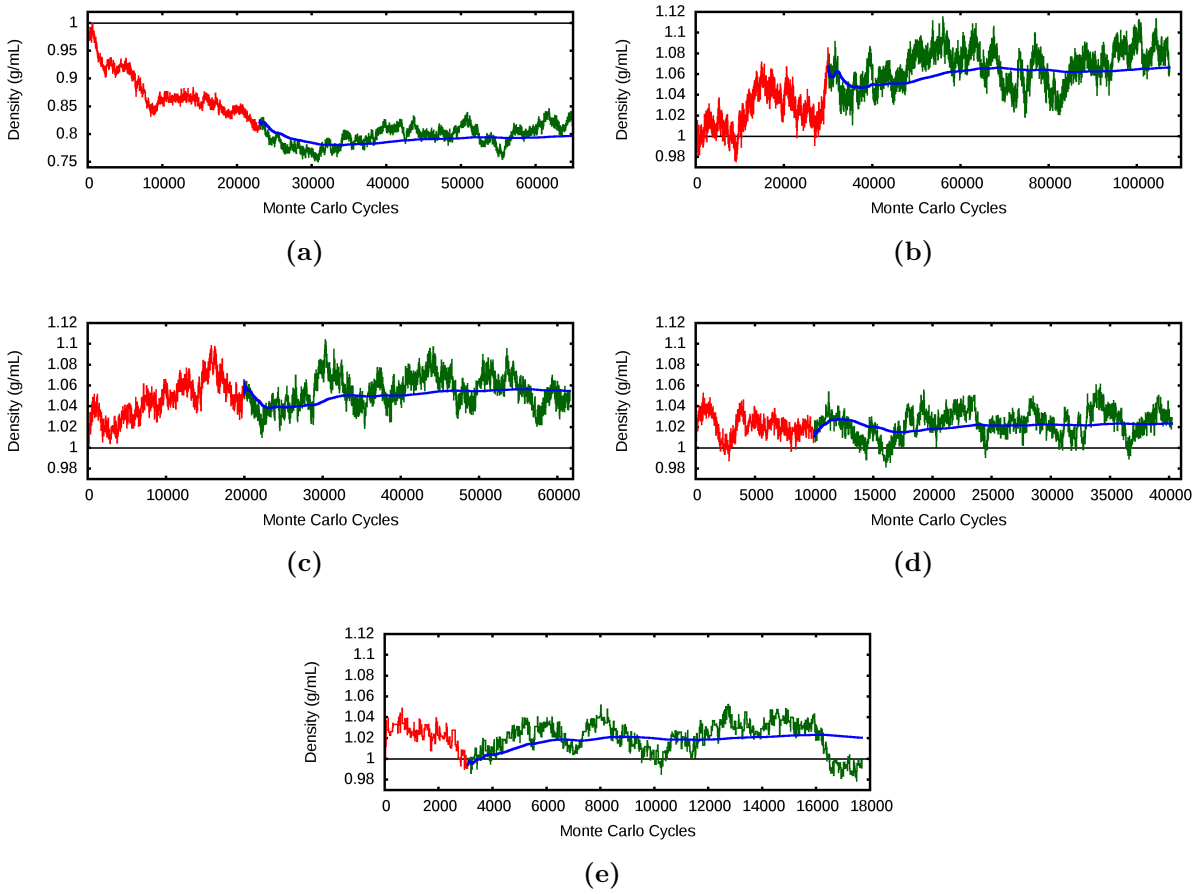


Figure 7.2: Fluctuation of the instantaneous density as a function of the Monte Carlo cycles for the NpT simulation at the various level of theory considered ($T = 295$K and $p = 1$bar). The red portion of the plots denotes the cycles considered for equilibration, the green parts refers to that used for the calculation of the average properties, the blue line shows the running average density. In each plot the horizontal black line represents the experimental density. Note the different ordinate scale in panel (a). (a) BLYP, (b) BLYP-D3, (c) PBE-D3, (d) PBE0-ADMM-D3, (e) RI-MP2.

contrast to that of BLYP, a dispersion free functional that yields 0.797 g/mL. The reason for this is that the water density depends crucially on the medium to long range part of the potential [276, 303]. Dispersion corrected density functionals, BLYP-D3, PBE-D3, and PBE0-ADMM-D3 all have densities that are much closer to experiment, albeit slightly too dense. The density obtained for PBE0-ADMM-D3 (1.023 g/mL) is of MP2 quality, but this might be in part fortuitous, as the small basis employed in the ADMM approach could influence this result. Nevertheless, it is known that the PBE0 functional improves PBE results for structural, spectroscopic and thermodynamics properties [173, 304, 305], and yields good results for water dimer interactions when compared to high level coupled cluster calculations [306].

Computing a first principles estimate of the density is challenging, and long simulations are essential to sample fluctuations and to equilibrate the system. If a significant structural reorganization would be needed, equilibration times could easily exceed simulation

	Density [g/mL]		1^{st} Max		1^{st} Min		2^{nd} Max		CN
	ρ	RMSD	r [Å]	$g_{OO}(r)$	r [Å]	$g_{OO}(r)$	r [Å]	$g_{OO}(r)$	n_{OO}
BLYP	0.797	0.018	2.83	3.04	3.46	0.44	4.65	1.37	3.9
BLYP-D3	1.066	0.018	2.78	2.78	3.51	0.92	4.37	1.11	5.9
PBE-D3	1.055	0.015	2.73	3.07	3.25	0.69	4.43	1.21	4.2
PBE0-ADMM-D3	1.023	0.013	2.74	3.13	3.30	0.65	4.52	1.25	4.4
RI-MP2	1.020	0.015	2.76	3.05	3.32	0.72	4.41	1.21	4.5
exp.	1.00	-	2.80	2.57	3.45	0.84	4.5	1.12	4.3

Table 7.1: Average density and structural data obtained from the MC NpT simulations at the various level of theory considered ($T = 295$K and $p = 1$bar). Experimental values taken from Ref. [297]. CN is the average coordination number calculated from the integral $4\pi\rho_n r^2 g_{OO}(r)$ up to the first minimum, where ρ_n is the average number density of Oxygen atoms.

times. In this respect, our RI-MP2 simulation is a 'best effort' simulation, with a length constrained by the high computational cost. The number of MC cycles is larger in the corresponding DFT simulations. However, two observations enhance our confidence in the computed MP2 value. First, the total number of *accepted* MC cycles is similar for all reported calculations. This can be attributed to the high quality of DFT presampling in the MP2 case, which results in a much higher acceptance rate ($\sim 50\%$) compared to the other simulations using presampling based on a classical force field ($15 - 17\%$). Second, those methods that remain close to the experimental density (RI-MP2 and PBE0-ADMM-D3) are likely, and observed, to require shorter equilibration periods.

Further support for the employed methodology, but also insight in the typical error bars, comes from a comparison with literature DFT results. At the BLYP level, the calculated density is 0.797 g/mL, in agreement with the results reported by McGrath and coworkers [175] (0.8 g/ml), obtained by NpT-MC, Schmidt *et al.* [176] (0.73-0.78 g/mL) calculated employing NpT-BOMD, and Wang *et al.* [303] (0.76-0.85 g/mL). In the latter case, the water density is inferred from pressure density curves obtained from series of NVT-MD simulations performed at different volumes. A larger deviation is observed when the BLYP density is compared to the value calculated by Ma and coworkers [278] (0.92 g/mL) from NpT-CPMD using a DVR basis set. In that work, the large deviation is attributed to the usage of a converged basis set that contributes to softening the structure and improving diffusivity [307]. However, even though the basis set used here is significantly larger than in Ref. [176] and [175], a strong basis set dependence of the density at the BLYP level is not observed. The value calculated at the BLYP-D3 level (1.066 g/mL) is in agreement with that reported by Ma *et al.* [278] (1.07 - 1.13 g/mL). When comparing our BLYP-D3 and PBE-D3 with BLYP-D2 and PBE-D2 results reported by Schmidt and coworkers [176] deviations of 7% and 11% are observed, respectively. These deviations have to be attributed to the use of the D3 correction instead of D2, or potentially the use of larger basis sets and simulation lengths employed here.

The more detailed structure of the liquid is summarized in Figure 7.3 by the radial distribution functions for Oxygen-Oxygen ($g_{OO}(r)$) and Oxygen-Hydrogen ($g_{OH}(r)$), and quantified in Table 7.1. The MP2 results are in good agreement with experiment, in particular the mid and long range parts. The first minimum is relatively shallow,

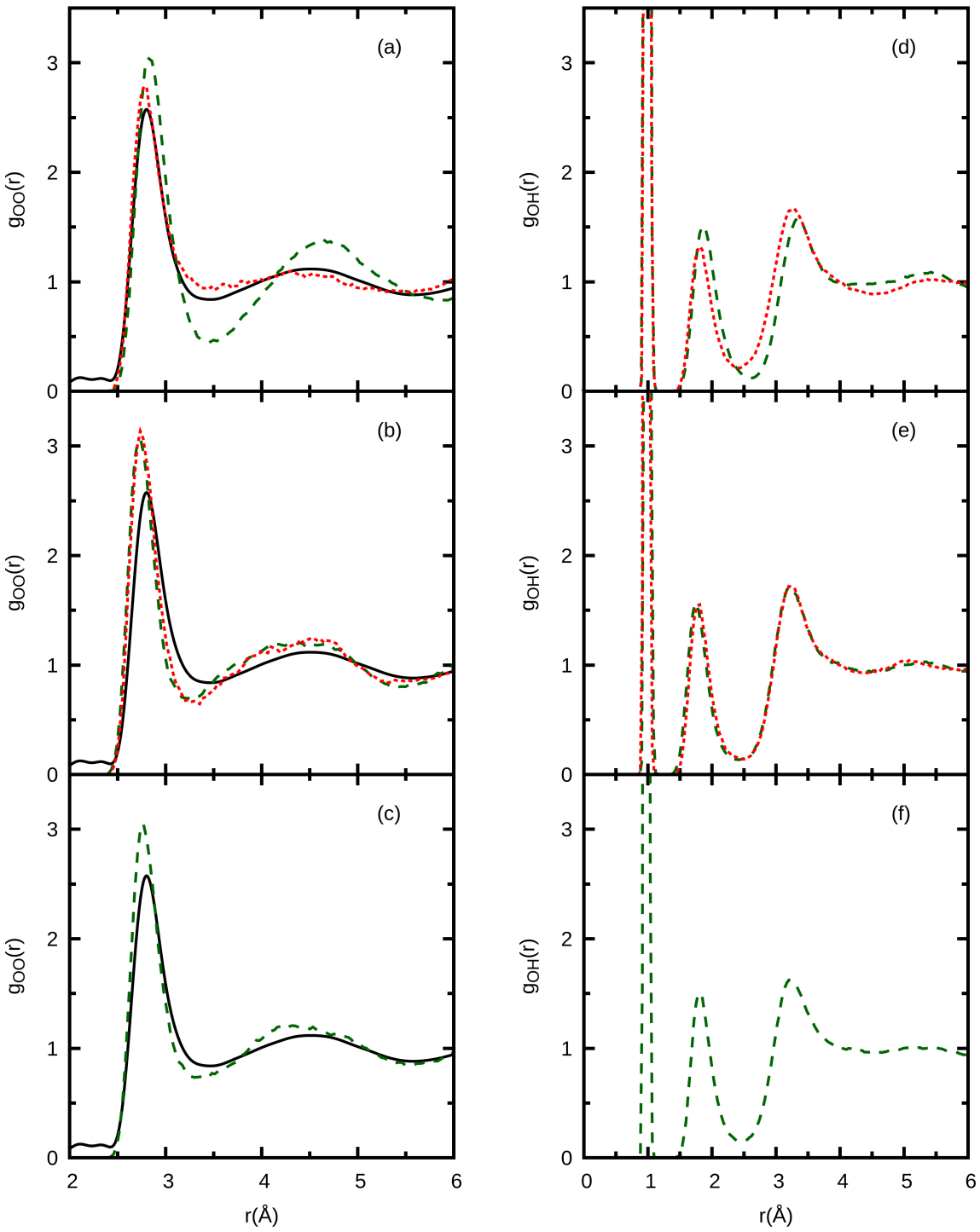


Figure 7.3: Radial distribution functions for Oxygen-Oxygen (left) and Oxygen-Hydrogen (right) distances obtained from the NpT-MC simulations at the various level of theory considered ($T = 295$K and $p = 1$bar). (a),(d) BLYP (green dashed) and BLYP-D3 (red dotted); (b),(e) PBE-D3 (green dashed) and PBE0-ADMM-D3 (red dotted); (c),(f) RI-MP2 (green dashed line). The most recent experimental Oxygen-Oxygen RDF, obtained from x-ray diffraction and taken from Ref. [297], is depicted as a solid black line.

suggesting that a diffusive liquid is obtained. However, MD is required to obtain a precise value for the diffusion constant, as time correlation functions can not be extracted from these MC simulations. The maximum of the first peak is too pronounced, i.e. 3.12 vs 2.57. This difference is in part explained by the fact that our simulations ignore nuclear quantum effects, which influence this property [298–301]. The coordination number, which condenses the shape of the first peak into a single number, is in fair agreement with experiment 4.6 vs. 4.3. The DFT results are less satisfactory, in particular BLYP without dispersion produces a very structured liquid as quantified by the low value of the first minimum. Note that the $g_{OO}(r)$ obtained under NpT conditions is significantly different from the one obtained in the NVE ensemble (see e.g. Ref. [203]) at experimental density. In particular, a low coordination number is observed. The D3 dispersion correction leads to a $g_{OO}(r)$ that is almost featureless after the first peak, such a smearing out of the second peak with inclusion of dispersion has been discussed by Møgelhøj et al. in Ref. [277]. The large coordination number can be considered an artefact of the shallow minimum observed, and depends strongly on the precise location of the minimum. PBE-D3 and PBE0-ADMM-D3 give very similar $g_{OO}(r)$, as was previously observed [104] in simulations without the dispersion correction. Whereas these functionals lead to a slightly more structured liquid, they are similar to the MP2 results, in particular PBE0-ADMM-D3. The latter functional could thus be a computational expedient alternative to MP2, for example to investigate nuclear quantum effects.

7.4 Conclusions

Results have been reported for liquid water at ambient conditions as obtained from NpT-MC simulations at the MP2 level theory, which is free from empirical parameters. The obtained density of 1.02 g/mL is in excellent agreement with the experimental value, and the $g_{OO}(r)$ generally agrees well with the most recent experimental observations, albeit with a slightly too high first maximum. Other popular DFT methods have been tested, in particular showing good results for the water density at the PBE0-ADMM-D3 level, while being generally in agreement with previously reported work for other DFT methods. It is clear that the inclusion of dispersion interactions and an accurate description of the hydrogen bond are crucial ingredients for the correct simulation of liquid water at ambient conditions. The quality and accuracy of the obtained MP2 results paves the way to further applications of this technique for challenging questions in the field of aqueous solutions, e.g. the structures of the solvated electron and hydroxyl ion.

Chapter 8

A periodic MP2, RPA and boundary condition assessment of hydrogen ordering in ice XV [308]

Ice XV is the hydrogen ordered form of the ice VI phase whose structure was predicted to be Cc and ferroelectric using periodic DFT approaches. However, neutron diffraction and Raman spectroscopy data show the structure to have $P\bar{1}$ symmetry and to be antiferroelectric. Recent work [309] using fragment-based MP2 and CCSD(T) approaches predict the experimental structure as the ground state. In this chapter this problem has been reconsidered using fully periodic MP2 and RPA approaches and it is found that the ferroelectric Cc structure is the lowest energy configuration. However, ubiquitously employed tin-foil boundary conditions stabilize polar structures. It is suggested that ferroelectric Cc crystals can grow within a paraelectric ice VI matrix, but may become unstable once a fraction of the matrix has become hydrogen ordered. The reduction in dielectric constant causes $P\bar{1}$ and other structures with small polarization to become favoured, providing a possible resolution between observation and theoretical predictions.

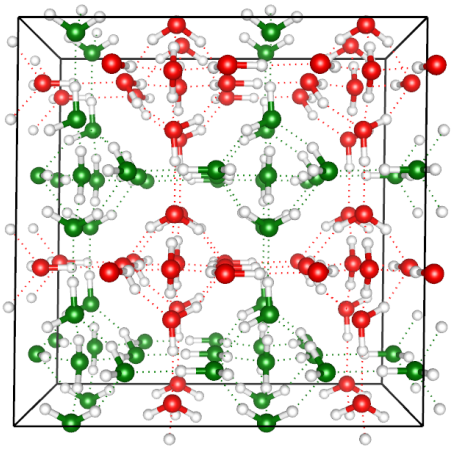
8.1 Introduction

Ice exhibits fifteen distinct crystalline phases [310] and the chemistry of the most readily accessible phase, ice Ih, continues to yield surprises, such as "cubic ice", which is more accurately described as stacking disordered ice Isd [311], a newly proposed ice 0 phase [312, 313] that appears to facilitate the homogenous nucleation of ice and the recent experimental identification of supercooled water beyond its accepted homogeneous freezing point [314]. High-pressure ice phases have also attracted attention and the most recently identified stable form of ice, ice XV [315] is a case in point. Almost a decade ago, Knight and Singer [316] and Kuo and Kuhs [317, 318] independently predicted that ice XV, a stable hydrogen disordered ice phase in the ≈ 1 GPa regime should have a crystallographically distinct hydrogen ordered structural form of Cc space group symmetry. Both groups predicted the same crystal structure, obtained using local density functional approaches, often referred to in the literature [310, 315] as the KSKK structure. When Salzmann *et al.* isolated and reported the hydrogen ordered form of ice VI in 2009, it

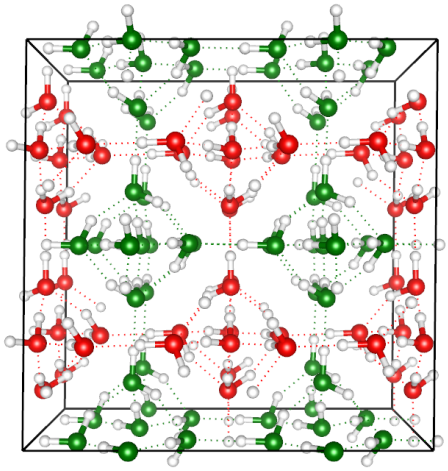
was found to have the $P\bar{1}$ crystal structure in contrast to the KSKK prediction. The discrepancy between theory and experiment was surprising in that numerous density functional theory studies at various levels of sophistication, had successfully identified the hydrogen ordered counterparts of ice I (ice XI [319–321]), ice II [322], ice III (ice IX [322]), ice V (ice XII [323]), and ice VII (ice VIII [319, 324, 325]). See also the review by Singer and Knight [326]. The mismatch between theory and experiment was highlighted in a recent review [310] where ice XV was proposed as a solid state benchmark system for hydrogen bonding. Recently Nanda and Beran [309] rose to the challenge and by using a hierarchical, fragment-based quantum mechanical/molecular mechanics approach, they found that the experimentally identified structure was favored over other structures at the MP2 and CCSD(T) levels, apparently resolving the conflict between experiment and theory. Aside from the crystallographic differences and corresponding electrical properties of each structure, another point of interest is that ice XV has the largest variation in hydrogen bonding angles of any crystalline ice phase. The latter property suggests that ice XV is a testing material for theoretical approaches; modeling of hydrogen bonds under pressure suggests that currently no density functional exists which can consistently predict density and the cohesive energies of phases with an accuracy comparable to diffusion Monte Carlo results [327, 328]. Moreover, in ice XV there is a coupled problem of describing the highly distorted arrangement of hydrogen bonds within each sub-lattice and the separation between the independent sub-lattices, which is affected by van der Waals interactions between the sub-lattices. van der Waals interactions are currently an area of very active development [329, 330] in the density functional field and in the absence of functionals that can provide chemical accuracy for solid cohesive energies over a wide range of densities, there has been progress in implementing embedding techniques [331, 332], periodic HF with CCSD corrections [333], periodic CCSD [334], local periodic MP2 methods [335, 336] and random phase approximation (RPA) approaches [183]. This motivated us to examine the ice XV problem using fully periodic RPA and MP2 treatments that have been recently implemented [44, 131] in the CP2K/Quickstep code [90, 91]. Although several studies of ice phases [183, 335] have been reported at the periodic MP2 and RPA levels, this is the first periodic study of ice XV at the MP2 and RPA level.

8.2 The Structure of Ice XV

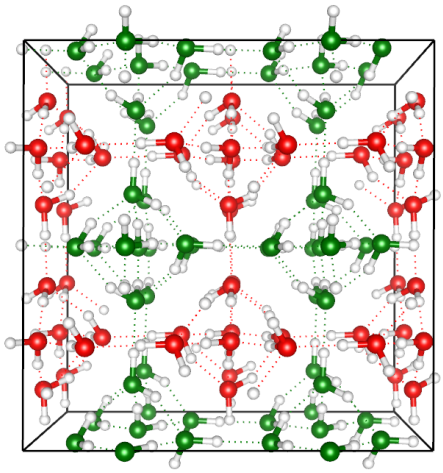
Figure 8.1a and Figure 8.1b show respectively the experimentally identified structure (denoted as SRMF) and the structure predicted by Knight and Singer and Kuo and Kuhs (KSKK). There is an important difference in the two lattices, the SRMF structure is anti-ferroelectric and the KSKK structure is ferroelectric. The structure consists of two interpenetrating lattices with the Edingtonite topology [337]; in the SRMF structure each sub-lattice is polar (and maximally so [310]) but the inversion symmetry of the $P\bar{1}$ space group means that the dipole moments of each sub-lattice cancel. In the KSKK structure each sub-lattice is also polar and the dipole moments of each sub-lattice only partially cancel giving rise to an overall net dipole for the unit cell. Using the oxygen net of ice VI, an individual network can hydrogen order in three different ways which have been labeled A, B and C, in accord with the original report of ice XV [315]. The unit cell contains 10 water molecules and according to symmetry analysis (using graph invariant software



(a)



(b)



(c)

Figure 8.1: Three of the possible eighteen structures for ice XV. Hydrogen is shown white spheres. Oxygen atoms are shown in green and red in order to distinguish the distinct sub-lattices within the structure. In each case a 2 x 2 x 2 supercell is shown along the [001] axis. (a) The experimentally determined SRMF ice XV structure (2C1) exhibiting $P\bar{1}$ symmetry. (b) The KSKK (9A2) structure predicted by first principles approaches exhibiting Cc symmetry. (c) The other Cc symmetry structure (9A1) which has a distinct hydrogen bonding pattern and sub-lattice orientation from the KSKK structure.

[338]), there are 18 symmetry inequivalent structures within the space groups 2, 4, 7 and 9 (corresponding to $P\bar{1}$, $P2_1$, Pn and Cc respectively) [310]. Note that there are two non-equivalent structures in the Cc (9) space group, which are labeled 9A1 and 9A2 after Ref. [309]. According to the analysis reported here, 9A2 (see Figure 8.1b) is isostructural with the KSKK structure, whereas Ref. [309] reports that the 9A1 (see Figure 8.1c) structure is the KSKK structure. 9A1 can be transformed to 9A2 by a 180 degree rotation of one of its sub-lattices about the c axis. The structural difference between 9A1 and 9A2 has a profound effect on their polarity as will be seen later. The figures of all possible 18

symmetry inequivalent structures of ice XV are reported in Appendix C.

8.3 Results

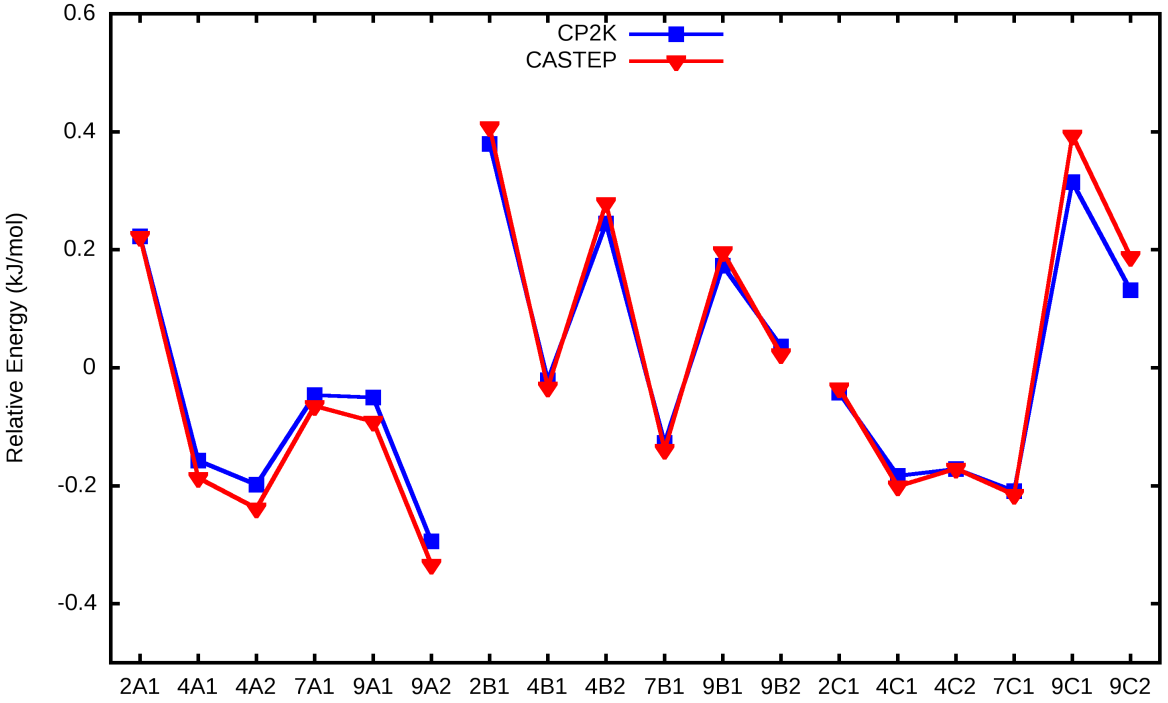


Figure 8.2: Relative energies, with respect to the average, of the 18 symmetry inequivalent structures possible in the ice XV unit cell reported at the PBE level of theory, consistently optimized with both codes. Energies on the vertical axis are expressed in kJ/mol relative to the average energy of all 18 structures and for the unit cell containing 10 molecules. Each structure type is labelled after the convention of Ref. [309].

The detailed description of the computational setups employed here are reported in Appendix C. As a consistency check for the computational methodology, two different computer programs have been used for performing a full cell optimization (without symmetry constraints, i.e. $P1$ symmetry) on all 18 of the symmetry inequivalent structures at the GGA PBE [170] level of theory (which yields reliable structures, see e.g. [192]). In Figure 8.2, the total energies are reported relative to the average energy computed using CP2K, which uses a dual local orbital and plane wave based scheme, and CASTEP [339], which uses different pseudopotentials and also a plane wave basis and is hence basis set superposition error free. More detailed study of basis set convergence, are described in Appendix C. The largest difference in energy was found to be ≈ 0.03 kJ/mol and suggests that the two computational procedures have similar accuracy, and also validates the use of atom centred orbitals, even for these small relative energies. In agreement with previous works [316–318] it was found that the global minimum is the KSKK structure (Cc space group), which is structure 9A2 in Figure 8.2. The experimentally identified structure is 2C1 and is unambiguously metastable with respect to several A, B and C type structures.

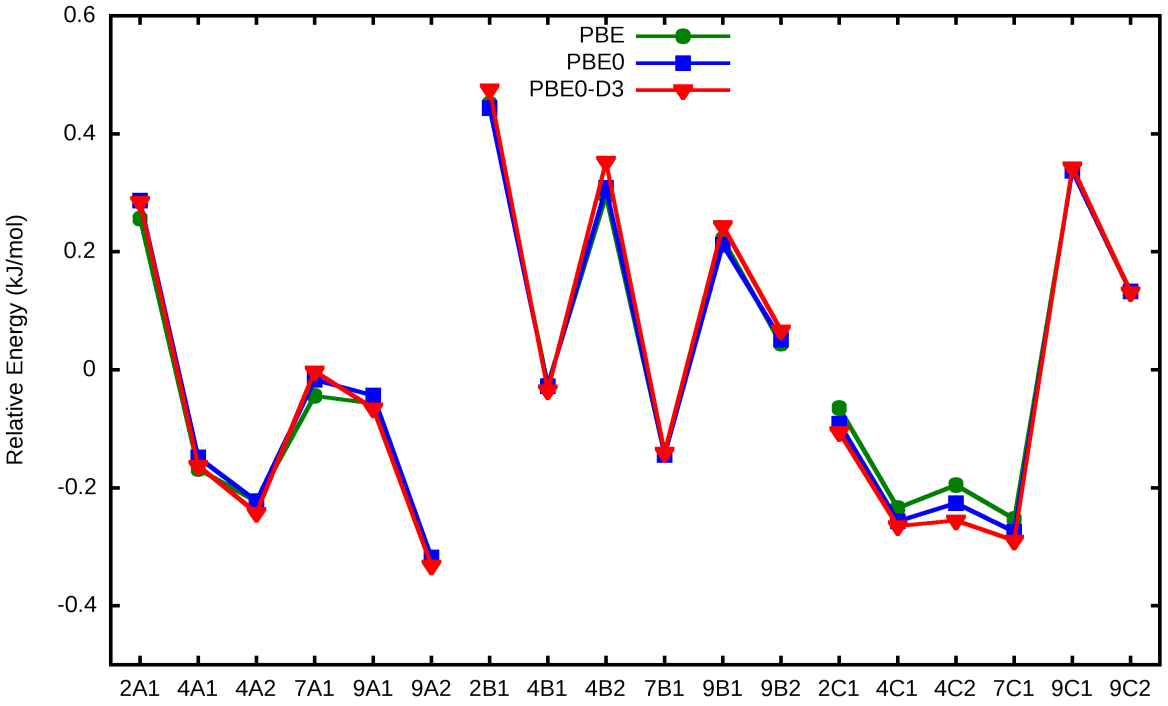


Figure 8.3: Relative energies, with respect to the average, of the 18 symmetry inequivalent structures possible in the ice XV unit cell reported at the PBE, PBE0 and PBE0-D3 level of theory employing MP2 optimized geometries.

To investigate if variations in exchange and correlation functional have a significant impact, the predictions from the hybrid PBE0 functional [173], with 25% HF exchange used in place of PBE exchange, as well as what the PBE0-D3 scheme have been examined. In the former scheme, the introduction of HF exchange opens the band gap, which is, for example, important to estimate the static dielectric constant of ice Ih [204]. In the latter scheme, D3 indicates the use of Grimme's van der Waals correction [124], which takes into account interactions that are missing from GGA density functionals in a semi-empirical way. The results are shown in Figure 8.3 and are remarkably consistent with the PBE results, and in particular, the 9A2 structure is again found to be the global minimum. Although relative energies of all structures are insensitive to the functional scheme employed, there is some variation in the absolute density. PBE gives 1.40 g/cm^3, PBE0 1.40-1.41 g/cm^3 and PBE0-D3 1.47-1.48 g/cm^3, in accord with similar studies of ice phases [327, 328] and liquid water [176, 177, 256]. Given the wide variations in intramolecular bond angle, variation in intermolecular hydrogen bonding angle and non-equivalent van der Waals interaction between the sub-lattices, the relative energies of the 18 possible structures could be sensitive to the precise formulation of the density functional employed. However, it is found that this is not the case for ice XV. In Appendix C more data are presented showing that this holds more broadly. Summarizing the results reported in Appendix C, other approaches for incorporating van der Waals interaction, based on a non-local kernel and the electron density, such as vdw-DF2 [340], optPBE-B88vdw [341] and rVV10 [342] also show very similar relative energies and also give 9A2 as the minimum energy structure. Also changing from the PBE family of functionals to the M06 family [343] of functionals,

with and without dispersion corrections, and including M06-2X-D3 scheme which uses 54% of HF exchange, leads to very similar results. A similar observation has been made previously for I/XI [344], which has a much smaller variance in the intermolecular hydrogen bonding angles.

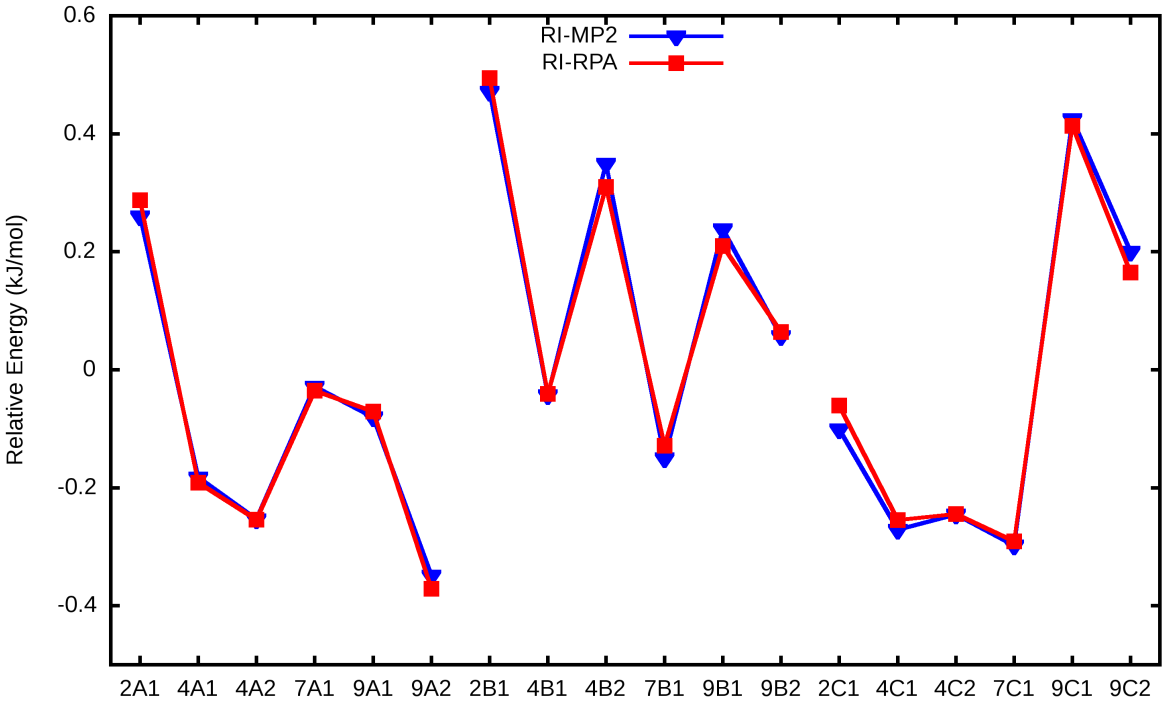


Figure 8.4: Relative energies, with respect to the average, of the 18 symmetry inequivalent structures possible in the ice XV unit cell reported at the periodic RI-MP2 and RI-RPA level of theory using the cc-QZ basis employing MP2 optimized geometries.

Higher level methods including wavefunction correlation offer an independent way of assessing the veracity of the density functional predictions. Nanda and Beran's work based on MP2 and CCSD(T), using a fragment-based, hybrid many-body interaction, quantum mechanical/molecular mechanics method, energetically favors the experimentally resolved $P\bar{1}$ (2C1) structure over the 2A1, 2B1, 9A1 and 9B2 structures. At the MP2 level, the remaining 13 structures were found to be significantly less stable. Here, all 18 structures have also been examined using fully *periodic* MP2 and RPA approaches. RPA can be viewed as an approximate CCSD method which only retains part of the diagrams [34]. RPA has been shown to yield relative energies for ice phases that approach the accuracy of quantum Monte Carlo methods [183]. For the MP2 method, forces and stresses have been implemented in CP2K [211], allowing fully consistent energies and geometries for a high level of theory. Basis set convergence of the relative energies is demonstrated in Appendix C, using basis sets up to the cc-5Z level (16000 basis functions for the simulation cell considered). Comparing MP2 single point energies for the structures obtained with the PBE0, PBE0-D3 and PBE approaches, we find that PBE0-D3 yields the lowest MP2 energy, making this a good starting point for subsequent optimizations. This observation is in agreement with previous work on liquid water, where it was found that that PBE0-D3 closely reproduces the MP2 density [177, 256]. The relative energies at the MP2 and RPA

level of MP2 optimized structures are displayed in Figure 8.4. We observe that MP2 and RPA relative energies are nearly identical and that they confirm the DFT results, with correlation factors larger than 0.995 for PBE, PBE0 and PBE0-D3 with the MP2 data. In agreement with Ref. [309], structures 2A1, 2B1, 9A1 and 9B2 all lie higher in energy than the experimental (2C1) structure but it is found that the KSKK structure (9A2) structure is the global minimum. At all levels of theory, the range of energies is ≈ 0.8 kJ/mol per molecule in good agreement with the ≈ 0.8 kJ/mol per molecule found in Ref. [316] (periodic local density functional approach) and ≈ 1 kJ/mol per molecule found in Ref. [322] (periodic semi-local density functional approach). This range is substantially less than that of >15 kJ/mol per molecule reported in Ref. [309]. We also note non-negligible individual discrepancies between the fully periodic MP2 energy differences and those reported in Ref. [309] using fragment-based approaches. For example, we find that the 2C1 structure is more stable than 9B2 by 0.1 kJ/mol whereas in Ref. [309] an energy difference of ≈ 0.8 kJ/mol is reported. The uniformity of the energy differences between structures obtained here, with a range of periodic electronic structure approaches and with two independent codes (CASTEP and CP2K), gives us confidence in the order of the reported stability and absolute energy differences. As an additional check on the sensitivity of the results to the starting geometry used, in Appendix C results are reported at the MP2 and RPA level, obtained by using the geometries given in Ref. [309] and those obtained here and no significant discrepancies have been found. The density of MP2 ice XV is found to be 1.49 g/cm^3, closely matching the density predicted by the PBE0-D3 results but significantly greater than the experimentally inferred value of 1.328 g/cm^3 (1.476 g/cm^3 for the experimentally isolated deuterated form). Looking closely at the relative energies in Figure 8.4, whilst it is clear that the experimental 2C1 structure is metastable with respect to 9A2, structures 4A2, 4C1, 4C2 and in particular 7C1 are quite comparable to 9A2. One possibility is that the experimental structure is a product of the slow kinetics of transformation and that annealing of the experimental sample could lead to other observed structures, such as 4C1 and 7C1. However, occupancies of the hydrogen crystallographic positions that are unambiguously associated with 2C1 increased upon cycling the sample above and below the transition temperature regime. Subsequently, further analysis, including Raman assessment [345], has shown that only the 2C1 structure can account for the diffraction and spectroscopic signatures.

8.4 Introducing Dielectric Boundary Conditions

Since there is a clear mismatch between what is seen experimentally and what is consistently predicted by all levels of theory considered here, a possible explanation for the discrepancy is proposed. In previous studies of hydrogen order in ice, in particular ice XI [346] and also the work of Nanda and Beran, the importance of convergence of long-range interactions has been highlighted. These electrostatic interactions are computed with an Ewald sum (see Allen and Tildesley [347]) or related approaches. In the common definition and implementation of the Ewald scheme, so-called tin-foil or metallic boundary conditions are employed, equivalent to assuming that the repeated unit cell is embedded in spherical cavity of a medium with infinite dielectric constant (e.g. a metal). The question here is if these metallic boundary conditions are appropriate to study the ice VI

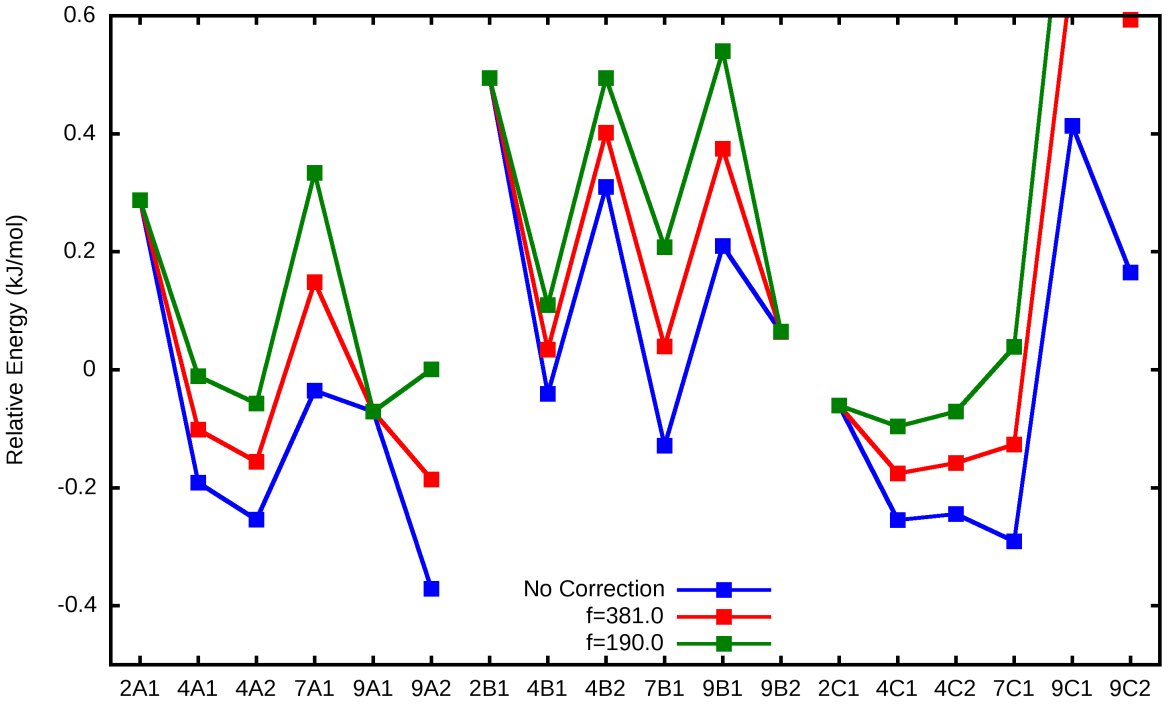


Figure 8.5: Relative energies, with respect to the average, at the RI-RPA level of theory for the 18 symmetry inequivalent structures in the ice XV unit cell, including corrections for different dielectric boundary conditions.

to ice XV transformation. Whilst the influence of electrical boundary conditions has been considered in the field of classical modelling, their effects in periodic quantum mechanical calculations are comparatively rarely considered. A notable exception is recent work on ferroelectric solids, such as the work by Stengel *et al.* [348] (and references therein), where an elegant formalism has been proposed to explore the coupling between internal energy and the applied electric field in a consistent manner. The hydrogen disordered ice VI phase has a high dielectric constant, estimated by computation (after correction) to be 181 at 243K according to Ref. [349], consistent with the only reported experimental value of 176 taken from Ref. [350]. This value increases with decreasing temperature, as long as the phase remains hydrogen disordered. On the other hand, ordered ice phases have low dielectric constants, around 3-5, depending on temperature. Hence, at nucleation, ice XV crystals are effectively embedded in a medium with relatively high but finite dielectric constant, and the tin-foil approximation would be reasonable. However, as a larger fraction of the ice VI transforms into an ordered phase, the effective dielectric of the medium would be reduced, and nucleation and growth could take place in a matrix of effective low dielectric constant. In line with this model, Johari and Whalley [351] reported evidence of ferroelectric ordering, and a reduction in the dielectric constant of a ice VI sample with time, monitored over 252 days at a constant 117 K. However, the molecular structure of the ordered ice VI was not analyzed.

To model the relative stability of ice XV nucleating within a medium of low dielectric constant, the Ewald construction requires an additional surface correction term, detailed

within the works of de Leeuw *et al.* [352] and Hummer *et al.* [353]:

$$U_{corr} = \frac{2\pi}{Vf} M^2. \tag{8.1}$$

Equation (8.1) gives an *a-posteriori* correction to the total energy, where M is the dipole moment of the simulation cell of volume V, and f is $(2\epsilon+1)$, when the system is embedded in a spherical cavity of a medium of dielectric constant ϵ. Note that the precise expression for f depends on the shape of the cavity, and will be different for platelet or needle shaped crystals. This term vanishes if the system has no net polarization or if the dielectric constant of the medium becomes large, but will destabilize ferro-electric systems if the dielectric constant of the medium is small. To test the sensitivity of the predicted total energies of putative ice XV nuclei to the presence of a medium with finite dielectric constant, the polarization as obtained for a hybrid density functional calculation on top of the MP2 geometries have been employed adding the correction (Equation (8.1)) to the computed RPA energies for $f = 0$, $f = 190$ and $f = 381$, these results are reported in Figure 8.5. Note that, given the large uncertainty in the dielectric constant and unknown shape of the crystallites, the precise value of f is of secondary importance. However, it is immediately clear that there is a substantial influence on the relative stability of ice nuclei depending on the dielectric properties of the surrounding medium. In particular, the most striking point is that the polar structures including 9A2 are substantially destabilized by the presence of a medium with finite dielectric constant, while the structures with zero (2A1, 2B1, 2C1) and small (9A1 and 9B2) dipole moment remain unaffected. As the dielectric constant of the ice VI medium diminishes, 2A1, 2B1, 2C1 and 9A1, 9B2 become the most stable hydrogen arrangements. Ultimately this yields the experimental structure 2C1 essentially iso-energetic with 9A1, more stable by 0.02 kJ/mol at the MP2 level, less stable by 0.01 kJ/mol at the RPA level, a difference that is too small to resolve by current computational procedures. However, at intermediate values of f, the 4A2, 4C1, 4C2 and 7C1 structures appear likely. As shown schematically in Figure 8.6, here it is speculated that small nuclei of ferroelectric 9A2, KSKK crystallites form within ice VI but antiferroelectric crystallites become favored as the dielectric constant of the transforming ice VI reduces. It could be evidence for this model that experimentally, only partial conversion from ice VI to ice XV is achieved of around 78%, indicating that a proportion of the matrix is unable to transform, possibly due to the presence of ferroelectric ice XV kinetically trapped in paraelectric ice VI and antiferroelectric ice XV. Finally, other factors could also influence which hydrogen arrangement is favoured. For example, DCl, which is used as a dopant to facilitate the phase transition, could influence the resultant phase, but the dopant levels used in the isolation of XV are rather low, approximately 1 DCl for 5500 D_2O molecules.

8.5 Conclusions

To summarize the main findings:

1. It has been established that high-level DFT and periodic MP2 and RPA methods yield a ferroelectric Cc hydrogen ordered structure as the lowest energy structure in

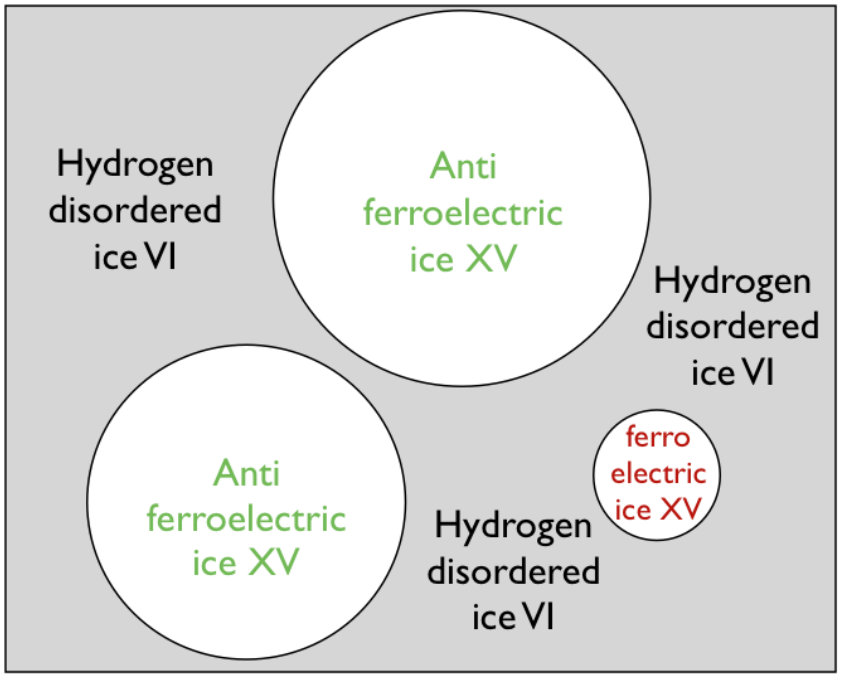


Figure 8.6: Schematic of the emergence of small ferroelectric KSKK nano crystallites, whose energy depends on the dielectric properties of the surrounding mixed ice VI-XV matrix. Anti-ferroelectric crystals however, can form without penalty.

agreement with early predictions [316–318] but in disagreement with more recent theoretical work [309] and the experimentally observed $P\bar{1}$ structure.

2. A possible explanation for the disagreement between theory and experiment is proposed by invoking a simple model of the electrostatic environment in which ice XV crystals grow.

This model suggests that complete transformation to ferroelectric ice XV is disfavored under open-circuit conditions but might be favored under closed circuit conditions and *e.g.* in an applied electric field. The plausibility of (2) could be tested directly by experiment, and indeed Jackson and Whitworth [354] have reported the influence of electric fields on the formation of partially ordered ice Ih and reported evidence of the formation of a ferroelectric structure, but it's detailed atomic structure was not resolved. The same mechanism could also occur in the ice Ih/XI transition and could help to explain why ice XI has never been made above $\approx 70\%$ pure (in the absence of an applied field). Finally, consideration of the electrical boundary conditions might be generally important in the field of crystal structure prediction.

Chapter 9

Summary and Outlook

Summary

The whole work reported in this thesis has as a common focus the introduction and application of methods that allow for accounting for non-local dynamical electron correlation. Specifically, this has been accomplished via many-body perturbation theory, in the Møller-Plesset formulation (MP2), and random phase approximation (RPA) correlation energy method. Both of them represent well known approaches since the early times of quantum mechanics, but while the former has been extensively used, for the latter there has been a revived interest only recently thanks to the introduction of low order scaling algorithms for its treatment. An additional difference is that MP2 is much more popular in the chemistry community and thus used to study molecular (gas phase) systems, while RPA has been developed and more commonly employed for condensed phase matter applications. An overview of the theoretical framework for MP2 and RPA is given in Chapter 2.

In this respect, the earlier stages of this work have been focused on extending the MP2 method to condensed phase systems. This has been accomplished by applying the well established Gaussian and Plane-Waves (GPW) approach in order to deal with mixed occupied-virtual electrostatic densities (Chapter 3). The inclusion of the correct boundary conditions for treating systems in the condensed phase had as a drawback that the computation of the needed matrix elements is much more expensive than the relative task in molecular cases. On the other hand the developed scheme had an easy and straightforward way for parallel implementation and allowed for the design of an algorithm capable, at that time, to scale up to the world largest supercomputers. In this way, for the first time, it was possible to calculate MP2 energy of condensed phase systems containing hundreds of atoms and thousands of basis function within minutes. Nevertheless, a detailed analysis of the algorithm, showed that the integral computation part, even if with a favorable system size scaling, was dominating the overall time to solution also for the larger applications.

The main achievement and breakthrough of this thesis work is represented by the combination of the resolution of the identity (RI) approximation with the previously introduced GPW method (Chapter 4). The RI approximation is essentially a way to factorize higher rank tensors by lower order ones. In the specific case, this is accomplished with a fitting procedure of the mixed occupied-virtual electrostatic densities in terms of a

set of auxiliary basis function. The main advantages related to the RI approximation is a reduction in both computational and memory requirements associated with the integral computation part of the algorithm. In this respect the developed RI-GPW method speeds up the integral computation by order of magnitude compared to the canonical GPW approach, while keeping the parallel performance essentially unchanged. Additionally within the RI approach the most intense part of the calculation are formulated in terms of matrix multiplication, that represents a very highly optimized task in high performance computing. The RI-GPW method has been applied to the calculation of the MP2 and RPA energy. In the latter case the combination of the RI approximation with imaginary frequency integration allows to formulate the RPA energy computation with a low order scaling procedure opening the possibility to treat systems that were not feasible before.

The developed RI-GPW method implemented with an algorithm suitable for massively parallel computation allowed to perform the first wavefunction based *ab initio* simulation of a complex condensed phase system, namely bulk liquid water (Chapter 7). The possibility to treat this problem from an high level quantum mechanical approach together with the quality of the obtained results is encouraging for further applications. In this respect, the study of the stability of the different proposed structures of ice XV represents just one of these possibilities (Chapter 8).

Having accurate total energies is of prime importance in quantum chemistry since the properties of a system, that can be related with experimental observables, are basically calculated as derivatives of the energy. Thus, once the machinery for computing the correlation energy is well established, a natural way to proceed is to calculate properties at the same level of theory in terms of derivatives. As an example, the forces acting on the atoms and the stress tensor are crucial for gaining insights on the structural properties of systems in the condensed phase as well as basic ingredients needed in molecular dynamics simulations (MD). These quantities are obtained as derivatives of the total energy with respect to the atomic coordinates for the former and cell parameters for the latter. The equations for evaluating the derivatives of the MP2 energy in the framework of the RI-GPW method have been derived and based on this a massively parallel algorithm has been developed for finite and extended system (Chapter 6).

As a final remark, the continuously increasing demand for computational power, has led to supercomputer facilities that are based more and more often on hybrid architectures. In an hybrid machine, different computing units, each of them particularly suitable for accomplishing a given set of operations, are combined together in order to enhance the overall performance. An example of these architectures is given by the combination of standard CPU with graphics processing unit (GPU). It is clear that the design of algorithms in the future will have to face this evolution in high performance computing and this is how the algorithms in this thesis work have been developed (Chapter 5).

Outlook

The work presented here represents only the basis for many possible follow up, for example in the direction of method development, high performance computing and applications.

The wide and continuously growing usage of the resolution of identity approximation in

computational chemistry and physics offers the possibility to extend the existing machinery to other treatment of electron correlation (*e.g.* Coupled Cluster). Also the RPA correlation energy method is a very active field of research, and many improved variants have been developed that can be readily implemented starting from the existing code. In this respect, the machinery for computing the MP2 energy derivatives shares many analogies with the recently formulated RPA energy derivatives, giving the possibility to include them in a relatively straightforward manner. Additionally many intermediate steps of the MP2 energy derivatives method are the same as those needed in the calculation of excitation energies and thus can be used as a basis for this purpose.

One of the flaws of the MP2 and RPA method is the slow convergence of the correlation energy with respect to the basis set size due to the electron coalescence cusp. A possible way to accelerate the convergence is offered by explicitly correlated schemes that can be extended in order to be used within the GPW approach.

To extend the applicability of the MP2 and RPA methods to larger systems it is of prime importance to developed algorithm suitable to exploit the continuously increasing power of the state-of-the-art high performance computing infrastructures. In this respect the adaptation of the existing code for being able to take full advantages from hybrid architectures is an ongoing project. Just as an example the improvements in the performance for the matrix multiplications between skinny matrices on GPU has already shown promising results.

As a final remark, it can be noted that, due to the relatively recent introduction of MP2 and RPA for the reliably study of condensed phase system, the possible applications are virtually infinite. In this respect these methods can be either used as reference results for benchmarking lower level methods, or as high level approaches when the inclusion of non-local dynamical correlation is essential.

Appendix A

Minimax Approximation

A fundamental problem in Approximation Theory is how an interpolating function (such as a polynomial) can be used to approximate a given target function [355]. This problem is referred as Fitting Problem and can be stated more conveniently in this context [179] as: given a function $f(x) : I = [1, R] \to \mathbb{R}$ with $R > 1$, and a model function $\phi(\alpha, x) : \mathbb{R} \times I \to \mathbb{R}$, find the set of parameters $\vec{\alpha} = (\alpha_1, \cdots, \alpha_N), \vec{\beta} = (\beta_1, \cdots, \beta_N) \in \mathbb{R}^N$, such that the error function

$$\eta(\vec{\alpha}, \vec{\beta}, x) = f(x) - \sum_{i=1}^{N} \beta_i \phi(\alpha_i, x) \tag{A.1}$$

is minimal with respect to a given norm. In general, different norms lead to different approximations, for example the minimization of $\eta(\vec{\alpha}, \vec{\beta}, x)$ with respect to the *2-norm*

$$\|\eta\|_2 = \int_1^R \left|\eta(\vec{\alpha}, \vec{\beta}, x)\right|^2 dx \tag{A.2}$$

gives the least squares approximation and associated least squares coefficients $\vec{\alpha}^0, \vec{\beta}^0$.

In the context of Minimax approximation [187] the minimization of $\eta(\vec{\alpha}, \vec{\beta}, x)$ is obtained with respect to the *infinity-norm*, defined as

$$\|\eta\|_\infty = \max\left\{\left|\eta(\vec{\alpha}, \vec{\beta}, x)\right| \in \mathbb{R} : x \in I\right\} \tag{A.3}$$

giving the associated Minimax coefficients $\vec{\alpha}^*, \vec{\beta}^*$. According to the alternation theorem of Chebyshev, there exists a unique set of $2N + 1$ extremum points

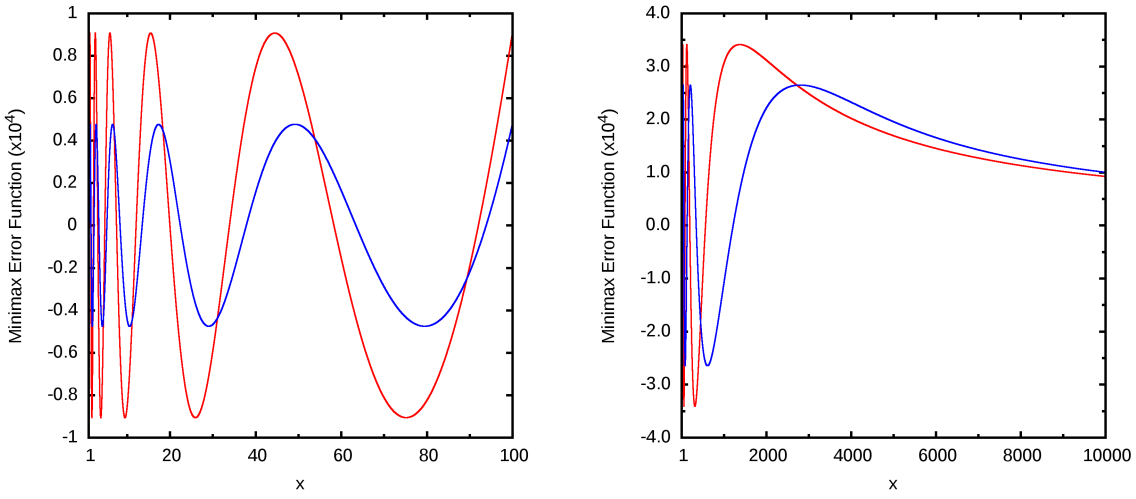
$$\{x_1 = 1 < x_2 < \cdots < x_{2N+1} = R\}$$

such that the error function satisfies

$$\eta(\vec{\alpha}^*, \vec{\beta}^*, x_j) = (-1)^j \, \epsilon \quad \forall \, j = 1, \cdots, 2N + 1 \tag{A.4}$$

where ϵ is the error extremum. According to Equation (A.4), there are $2N + 1$ linearly independent equations, $2N$ of them allow for optimizing the $\vec{\alpha}$ and $\vec{\beta}$ parameters, and one for minimizing the maximum error ϵ. The former $2N$ equations are obtained by eliminating ϵ in Equation (A.4)

$$\eta(\vec{\alpha}^*, \vec{\beta}^*, x_j) + \eta(\vec{\alpha}^*, \vec{\beta}^*, x_{j+1}) = 0 \quad \forall \, j = 1, \cdots, 2N. \tag{A.5}$$



(a) **(b)**

Figure A.1: Minimax error function $\eta(\vec{t}^*, \vec{w}^*, x)$ (blue) and $\eta(\vec{\nu}^*, \vec{\gamma}^*, x)$ (red) obtained with $N = 6$ for (a) $R = 100$ and (b) R exceeding the upper bound $R^{\max}$.

The importance of the fitting problem lies in the fact that its solutions can be used for finding optimal quadrature formulas in the numerical evaluation of integrals. Considering the evaluation of the MP2 energy in the Laplace-transformed technique [62, 63], the energy denominator is replaced by the associated integral in the time domain

$$\frac{1}{\epsilon_i + \epsilon_j - \epsilon_a - \epsilon_b} = \int_0^\infty e^{(\epsilon_i+\epsilon_j-\epsilon_a-\epsilon_b)t} dt, \tag{A.6}$$

that is carried out numerically by using the quadrature formula

$$\frac{1}{x} \simeq \sum_i^N \beta_i e^{-\alpha_i x}, \tag{A.7}$$

that ultimately represents a fitting problem for which the $1/x$ function has to be approximated by a sum of exponential functions over the range of definition of the energy denominator. A similar fitting problem has also been established for the frequency integral in the evaluation of the RPA correlation energy [179]. In this case, even if not in such a straightforward manner as in the Laplace-MP2 method, it has been shown that the interpolation function has the following form:

$$\frac{1}{x} \simeq \frac{1}{\pi} \sum_i^N \beta_i \left(\frac{2x}{\alpha_i^2 + x^2} \right)^2. \tag{A.8}$$

According to Equations (A.7) and (A.8) the two error functions employed in the fitting procedure read

$$\eta(\vec{t}, \vec{w}, x) = \frac{1}{x} - \sum_i^N w_i e^{-t_i x} \tag{A.9}$$

Table A.1: Values of $R^{\max}$ for $\eta(\vec{t},\vec{w},x)$ and $\eta(\vec{\nu},\vec{\gamma},x)$.

N	$\eta(\vec{t},\vec{w},x)$	$\eta(\vec{\nu},\vec{\gamma},x)$
1	8.667	5.888
2	$4.154 \cdot 10^1$	$2.527 \cdot 10^1$
3	$1.468 \cdot 10^2$	$8.303 \cdot 10^1$
4	$4.361 \cdot 10^2$	$2.332 \cdot 10^2$
5	$1.154 \cdot 10^3$	$5.893 \cdot 10^2$
6	$2.807 \cdot 10^3$	$1.377 \cdot 10^3$
7	$6.373 \cdot 10^3$	$3.029 \cdot 10^3$
8	$1.375 \cdot 10^4$	$6.345 \cdot 10^3$
9	$2.839 \cdot 10^4$	$1.276 \cdot 10^4$
10	$5.650 \cdot 10^4$	$2.479 \cdot 10^4$
11	$1.089 \cdot 10^5$	$4.675 \cdot 10^4$
12	$2.042 \cdot 10^5$	$8.591 \cdot 10^4$
13	$3.737 \cdot 10^5$	$1.543 \cdot 10^5$
14	$6.691 \cdot 10^5$	$2.715 \cdot 10^5$
15	$1.175 \cdot 10^6$	$4.689 \cdot 10^5$
16	$2.027 \cdot 10^6$	$7.966 \cdot 10^5$
17	$3.440 \cdot 10^6$	$1.333 \cdot 10^6$
18	$5.753 \cdot 10^6$	$2.199 \cdot 10^6$
19	$9.491 \cdot 10^6$	$3.581 \cdot 10^6$
20	$1.546 \cdot 10^7$	$5.763 \cdot 10^6$

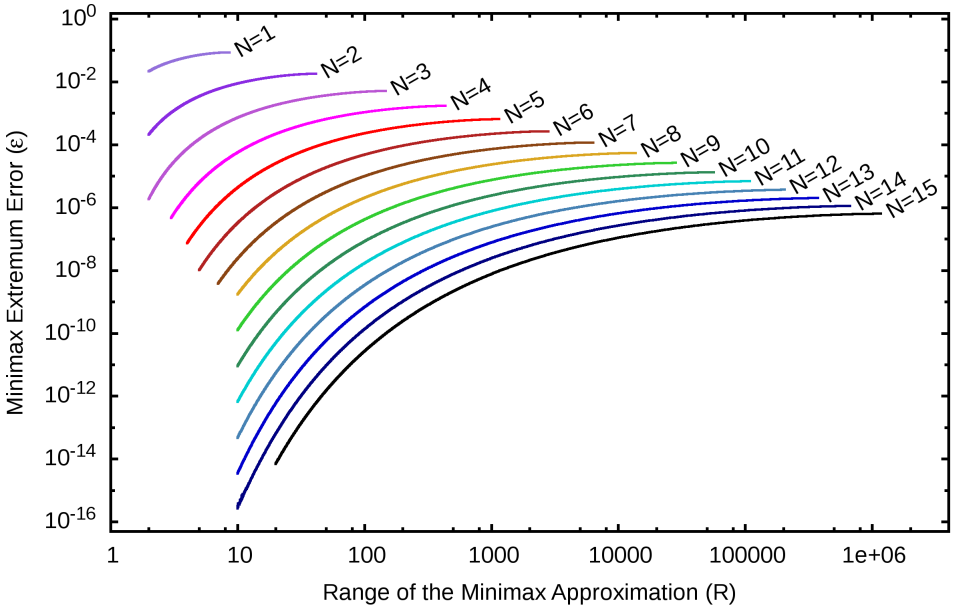
$$\eta(\vec{\nu},\vec{\gamma},x) = \frac{1}{x} - \frac{1}{\pi}\sum_{i}^{N} \gamma_i \left(\frac{2x}{\nu_i^2 + x^2}\right)^2 \tag{A.10}$$

for which the parameters $\vec{\alpha}$, $\vec{\beta}$ have been replaced with $\vec{t}$, $\vec{w}$ and $\vec{\nu}$, $\vec{\gamma}$ respectively for the MP2 and RPA case in order to distinguish between the time and frequency integration domains. The range of definition for these error functions is $x \in [1, R]$ with $R = E_{\max}/E_{\min}$, where $E_{\max}$ and $E_{\min}$ represent the maximum and minimum molecular orbital energy differences between occupied and virtual states respectively. This definition is more convenient for the optimization, and the final parameters employed for the evaluation of the numerical integrals are easily obtained by the following scaling relations:

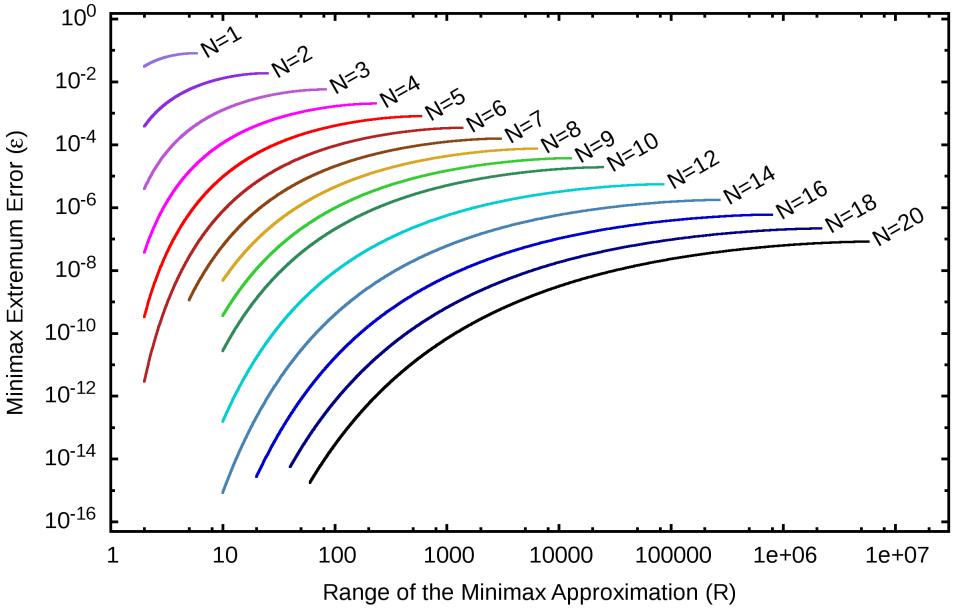
$$\bar{\nu}_i = \nu_i E_{\min}, \quad \bar{\gamma}_i = \gamma_i E_{\min} \tag{A.11}$$

$$\bar{t}_i = t_i/(2E_{\min}), \quad \bar{w}_i = w_i/(2E_{\min}). \tag{A.12}$$

With these quantities defined, the Minimax parameters $(\vec{t}^*, \vec{w}^*)$ and $(\vec{\nu}^*, \vec{\gamma}^*)$ are obtained by minimizing $\eta(\vec{t},\vec{w},x)$ and $\eta(\vec{\nu},\vec{\gamma},x)$ with respect to the *infinity-norm* $\|\eta\|_\infty$ making use of the alternation theorem [149, 150, 179]. This leads to the Remez algorithm:



(a)



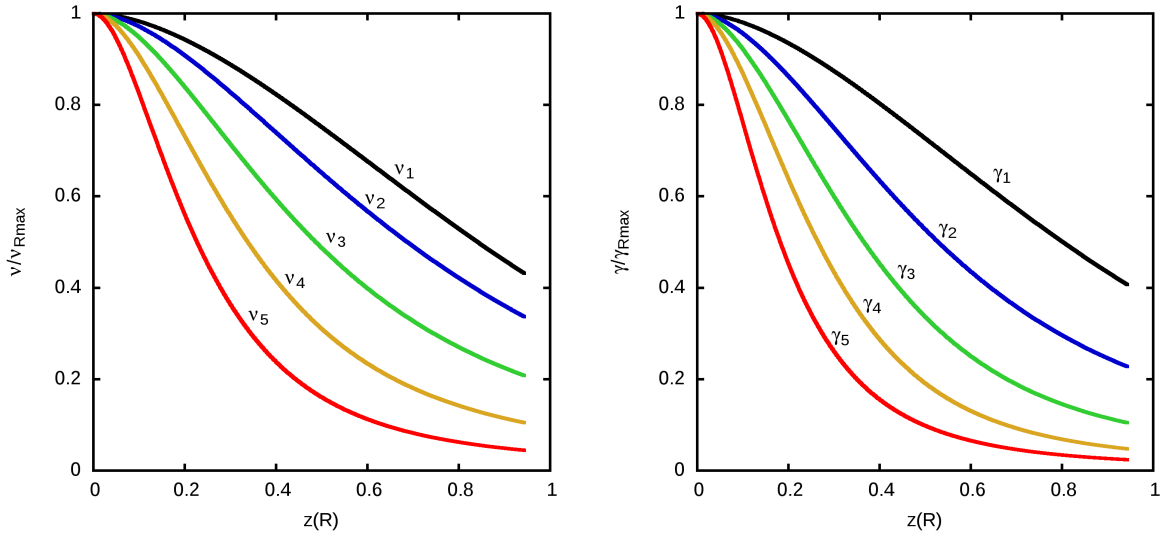
(b)

Figure A.2: Maximum error ϵ as a function of the approximation range R for different values of N. (a) $\eta(\vec{t}^*, \vec{w}^*, x)$ Equation (A.9), (b) $\eta(\vec{\nu}^*, \vec{\gamma}^*, x)$ Equation (A.10).

1. According to the value of R, generate an initial guess for the Minimax parameters from pretabulated data. The initial guess must form an alternant.

2. Find all extrema $\{x_j\}_{j=1}^{2N+1}$ of η.

3. Solve Equation (A.5) for the Minimax parameters, $(\vec{t}^*, \vec{w}^*)$ or $(\vec{\nu}^*, \vec{\gamma}^*)$, at the obtained extrema $\{x_j\}$.



(a) **(b)**

Figure A.3: Variation of the $\{\nu_i\}$ and $\{\gamma_i\}$ as a function of the remapped $R \to z(R)$, Equation (A.13). Each parameter has been divided by the relative one obtained for $R = R^{\max}$.

4. Update Minimax parameters.

5. Iterate till convergence is reached.

For the presented algorithm, a standard Newton-Raphson scheme can be used for solving the underlying nonlinear optimization problem. An example of Minimax optimized error functions ($\eta(\vec{t}, \vec{w}, x)$ and $\eta(\vec{\nu}, \vec{\gamma}, x)$) is shown Figure A.1a.

An additional detail that have to be considered in the Minimax approximation is that for any given N an upper bound $R^{\max}$ exists such that the associated error function $\eta(x)$, optimized for this intrinsic value, decays strictly in the semi-infinite interval $[R^{\max}, \infty)$. An example of this case is shown in Figure A.1b. Consequently, all Minimax approximations are identical for $R > R^{\max}$, the values of $R^{\max}$, for N up to 20, are reported in Table A.1.

In Figure A.2 the maximum error ϵ, obtained by optimizing $\eta(\vec{t}, \vec{w}, x)$ and $\eta(\vec{\nu}, \vec{\gamma}, x)$, is reported as a functions of R up to $R^{\max}$ for different values of N. It can be noted that, for a fixed value of R, increasing N leads to an exponential decay of ϵ, while, for each given N, the maximum error approaches the value associated to $R^{\max}$ by increasing the range of the approximation R.

The implementation of the Minimax approximation by mean of the Remez algorithm suffers essentially from two issues [155]. The first is that the algorithm requires an initial guess that should form an alternant, it turns out that if the quality of the starting guess is not good enough, the Minimax optimization could not be converged at all. The second is related to the fact that the algorithm, especially for large N, displays numerical problems even in double precision arithmetic.

In order to overcome these problems, the implementation of the Minimax approximation in CP2K [90] is based on a different strategy. For this purpose, the Remez algorithm has been implemented as a worksheet of the computer algebra system Maple [356] that allows

for performing multiple-precision floating-point computations. In this way Equation (A.5) can be converged with arbitrary precision, in particular, up to 100 significant digits have been employed for the larger N values and, in general, the convergence threshold have been set to 10^{-64}.

For each given value of N, a fine mesh of grid points for R has been generated in the range between a conveniently chosen $R^{\min}$ and the intrinsic value of $R^{\max}$. As previously mentioned, thanks to the properties of the Minimax approximation, it is not necessary to exceed $R^{\max}$, since all Minimax approximations are identical for $R > R^{\max}$. The Maple program has thus been used for optimizing the Minimax parameter for each R value defined over the grid. This implies that for each grid point $2N$ parameters are computed, N respectively for the $\{\alpha_i\}$ ($\{t_i\}$ or $\{\nu_i\}$) and $\{\beta_i\}$ ($\{w_i\}$ or $\{\gamma_i\}$).

Since the R grid is basically covering all possible Minimax approximations for each given N, the resulting parameters can be fitted, for example with a polynomial interpolation, offering the possibility to recover them without any need of further optimization. In order to do so efficiently, the values of $R \in [R^{\min}, R^{\max}]$ are mapped according to the following transformation:

$$z(R) = \left(\frac{R^{\min}}{R}\right)^{\frac{1}{4}} \left[\sqrt{\frac{R^{\max}}{R^{\max} - R^{\min}}} - \sqrt{\frac{R}{R^{\max} - R^{\min}}}\right] \tag{A.13}$$

$$z(R) \in \left[0, \frac{\sqrt{R^{\max}} - \sqrt{R^{\min}}}{\sqrt{R^{\max} - R^{\min}}}\right]. \tag{A.14}$$

The variation of the $\{\nu_i\}$ and $\{\gamma_i\}$ as a function of the remapped $R \to z(R)$, relative to the optimized parameters for $R = R^{\max}$, are shown in Figure A.3 in the case $N = 5$.

The resulting data are thus fitted over the new range of definition by using a 12^{th} order polynomial. In this respect, the fitted results are required to match a given target accuracy, if this is not the case, then the range is bisected and the fitting procedure is iterated till the required accuracy is fulfilled. The fitting procedure over a range is considered converged if the interpolating polynomial delivers parameters that display a maximum relative variation, with respect the "exact" ones, that is less than 10^{-9}. The Minimax parameters have been fitted up to $N = 15$ and $N = 20$ respectively for $\eta(\vec{t}^*, \vec{w}^*, x)$ and $\eta(\vec{\nu}^*, \vec{\gamma}^*, x)$.

Appendix B

RI Basis Optimization

One of the major components of the computational effort in many ab-initio methods in quantum chemistry is related to the computation of two-electron four-center electron repulsion integrals (ERI) over Gaussian basis or molecular orbital functions (see Equation (3.2)). A breakthrough in the reduction of the computational cost of this task has been achieved by the introduction of fitting procedures for approximating products of orbital functions by mean of linear combination of auxiliary Gaussian functions [357].

In general, the products of two orbital functions (either atomic orbitals or molecular orbitals) can be considered as an electronic density, and the fitting procedure can be cast as:

$$\rho_{ia}(\vec{r}) = \psi_i(\vec{r})\psi_a(\vec{r}) \approx \tilde{\rho}_{ia}(\vec{r}) = \sum_P C_P^{ia}\chi_P(\vec{r}) \tag{B.1}$$

in this case the i,a and P indices refer to occupied, virtual molecular orbital and auxiliary Gaussian basis function respectively. Unless the basis $\{\chi_P\}$ is spanning the whole space of the pair product $\psi_i \cdot \psi_a$, the approximation in Equation (B.1) introduces an error that has to be minimized. Minimization with respect of the Coulomb self interaction of the electronic densities $(\rho - \tilde{\rho})$, referred to as Coulomb metric [146], leads to an approximate factorization for the ERI's

$$(ia|jb) \approx (ia|jb)_{\mathrm{RI}} = \sum_{PQ}(ia|P)(P|Q)^{-1}(Q|jb) \tag{B.2}$$

in term of two and three center electron repulsion integrals. Equation (B.2) essentially resembles the insertion of the Resolution of the Identity (RI) operator and thus is referred as the RI approximation. Since the $(P|Q)$ matrix is positive definite, Equation (B.2) can be further factorized according to

$$(ia|jb)_{\mathrm{RI}} = \sum_Q B_Q^{ia} B_Q^{jb} \tag{B.3}$$

with

$$B_Q^{ia} = \sum_P (ia|P)(P|Q)^{-\frac{1}{2}} \tag{B.4}$$

giving the working equation for expressing the $(ia|jb)$ integrals in the RI approximation.

The fitting procedure outlined here is variational and the associated error is expected to decay rapidly with respect to the auxiliary basis set size due to the quadratic dependence

of the $(ia|jb)_{\mathrm{RI}}$ on the auxiliary expansion [145]. Nevertheless, in order for the RI approximation to be effective in reducing the computational cost, it is of prime importance to keep the number of auxiliary basis function as small as possible with respect to a given target accuracy. For this reason the generation of auxiliary basis sets that are specifically tailored on the employed primary basis represents an important step for the efficient application of the RI method. Here, the implementation of the RI basis optimizer in CP2K [90] is described with a particular focus on the technical details that have to be accounted to generate the auxiliary basis. The method is based on the work of Weigend and coworkers [69] that has been developed for RI-MP2 energy calculations. Nonetheless it has been shown that the usage of the same auxiliary basis for RI-RPA applications gives errors on the same order as those obtained at the RI-MP2 level [43, 133]. For completeness, the basic equations for the evaluation of the RI-MP2 energy in the general spin-orbital basis are also reported [68, 69].

For an unrestricted open-shell Hartree-Fock (UHF) reference the MP2 energy $E^{(2)}_{\mathrm{UHF}}$ reads

$$E^{(2)}_{\mathrm{UHF}} = \frac{1}{2} \sum_{\sigma_1\sigma_2} \sum_{ia\in\sigma_1} \sum_{jb\in\sigma_2} t^{ab}_{ij}(ia|jb) \tag{B.5}$$

with σ_1, σ_2 running over spins and the t^{ab}_{ij} amplitudes defined as:

$$t^{ab}_{ij} = \begin{cases} \dfrac{(ia|jb) - (ib|ja)}{\epsilon_i + \epsilon_j - \epsilon_a - \epsilon_b} & \text{if} \quad \sigma_1 = \sigma_2 \\ \dfrac{(ia|jb)}{\epsilon_i + \epsilon_j - \epsilon_a - \epsilon_b} & \text{if} \quad \sigma_1 \neq \sigma_2. \end{cases} \tag{B.6}$$

To improve readability, from now on the subscript UHF will be dropped. The application of the RI method to the computation of the MP2 energy ($E^{(2)}_{\mathrm{RI}}$) is straightforward and consists in the replacement of the $(ia|jb)$ integrals with the approximated $(ia|jb)_{\mathrm{RI}}$ given in Equation (B.3).

Due to the exchange integrals in Equation (B.6), the $E^{(2)}$ is neither an upper nor a lower bound of $E^{(2)}_{\mathrm{RI}}$. This implies that the minimization of the quantity

$$\Delta_{\mathrm{RI}} = E^{(2)}_{\mathrm{RI}} - E^{(2)} \tag{B.7}$$

with respect of the auxiliary basis parameters doesn't lead to a convenient optimization scheme. In order to overcame this complication Weigend *et al.* proposed to minimize the quantity ΔI defined as:

$$\Delta I = 2\tilde{E} - E^{(2)}_{\mathrm{RI}} - E^{(2)} \tag{B.8}$$

where

$$\tilde{E} = -\frac{1}{2} \sum_{\sigma_1\sigma_2} \sum_{ia\in\sigma_1} \sum_{jb\in\sigma_2} t^{ab}_{ij}(ia|jb)_{\mathrm{RI}} \tag{B.9}$$

with t^{ab}_{ij} defined in Equation (B.6) evaluated with non-RI integrals $(ia|jb)$. The quantity ΔI has the desired properties for the optimization procedure namely $\Delta I \geq 0$ and $\Delta I = 0$ if $E^{(2)}_{\mathrm{RI}} = E^{(2)}$. Moreover for a given converged wavefunction, ΔI is solely a function of the auxiliary basis functions.

For the actual implementation only non-contracted auxiliary basis are considered, meaning that only the exponents $\{\alpha_i\}$ of the Gaussian functions can be optimized. Thus, for a given initial guess made of N_a atom center Gaussian functions, the problem is to find the best set of exponents that fulfill

$$\frac{\partial \Delta I}{\partial \alpha_i} = 0 \quad \forall\, i = 1, \ldots N_a. \tag{B.10}$$

In order to do so, the derivatives in Equation (B.10) are calculated numerically and the optimization is performed with the Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm [358].

As a rule of thumb, in order for the RI approximation to be accurate enough, the RI-MP2 energy should display an error that is below 0.1 milli-Hartree per atom. Since ΔI can not be directly related to $|\Delta_{\mathrm{RI}}|$ the convergence is monitored by the quantity

$$\Delta I^{\mathrm{REL}} = \frac{\Delta I}{|E^{(2)}_{\mathrm{RI}}|}. \tag{B.11}$$

Even if the optimization of the RI basis can also be performed in the molecular case (with different atom kinds), it is usually performed for atoms. In this respect, following the work of Weigend *et al.* [69], the default convergence threshold for ΔI^{REL} is set to 10^{-6} and additionally $|\Delta_{\mathrm{RI}}|$ is required to be smaller than $50\,\mu E_h$.

The parallel algorithm for calculating ΔI is outlined in Figure B.1, for the actual implementation, the evaluation of the two and three center integrals is performed by using LIBINT, a library for the evaluation of molecular integrals over Gaussian functions [359]. Prior to the computation of ΔI, the non-RI MP2 energy and $(ia|jb)$ integrals have to be calculated. These quantities are independent of the RI basis parameters and thus are calculated only once at the beginning and stored in a replicated form over all N_p processes employed. The implementation of this step is based on the parallel direct canonical MP2 method [75–79].

The computation of ΔI and its derivatives is basically needed at each step of the optimization. Since the derivatives are computed numerically, this task simply implies the multiple computation of ΔI for which each of the exponent is slightly changed with respect to the actual value. Here, the parallelization strategy is different if the algorithm in Figure B.1 is used for computing ΔI or its derivatives. In the first case, all the N_p processes are used for the computation of ΔI, while in the second case the N_p processes are split into N_G groups of N_w workers ($N_p = N_G N_w$), each group working independently on a subset of preassigned derivatives. The size N_w can be given in input and the default value is $N_w = 1$. Figure B.2a shows the measured parallel speedup with respect to two processes for the computation of 20 steps in the RI basis optimization for a primary basis of Titanium at the triple-zeta level. In this case the number of exponents optimized is 33 and the employed group size is $N_w = 2$, more details on the calculation can be recovered from Table B.1. This implies that the algorithm displays an intrinsic falloff in the parallel performance for more than 66 processes, that corresponds to the case for which each group has only one derivative to compute. Thus the drop of the speedup in going from 48 to 64 processes is due to the imbalance of the work load over groups in the evaluation of the derivatives.

Assign each process its coordinate (n_p)

Create ranges $[P_{start}^{n_p}, P_{end}^{n_p}]$ over the N_a auxiliary function

Loop over P auxiliary basis functions ($P_{start}^{n_p} \leq P \leq P_{end}^{n_p}$)

Calculate $I_Q^P = (P|Q) = \int\int \chi_P(\vec{r}_1) \frac{1}{|\vec{r}_1 - \vec{r}_2|} \chi_Q(\vec{r}_2) d\vec{r}_1 d\vec{r}_2$ (all AUX Q)

End P Loop

Global summation of I_Q^P over all processes

Distribute I_Q^P into parallel distributed matrix $(P|Q)$

Cholesky decomposition of $(P|Q) = \mathbf{LL^T}$ (SCALAPACK)

Triangular inversion of $\mathbf{L} \rightarrow \mathbf{L^{-1}}$

Replicate $\mathbf{L^{-1}}$ into local buffer $(P|Q)^{-\frac{1}{2}}$ on each process

Loop over $\sigma = \alpha, \beta$

Loop over P auxiliary basis functions ($P_{start}^{n_p} \leq P \leq P_{end}^{n_p}$)

Calculate $S_{\mu\nu}^P = (\mu\nu|P) = \int\int \phi_\mu(\vec{r}_1)\phi_\nu(\vec{r}_1) \frac{1}{|\vec{r}_1 - \vec{r}_2|} \chi_P(\vec{r}_2) d\vec{r}_1 d\vec{r}_2$ (all μ, ν)

Atomic Orbital to Molecular Orbital transformation ${}^\sigma M_{ia}^P = \sum_{\mu\nu} {}^\sigma C_{\mu i} S_{\mu\nu}^P {}^\sigma C_{\nu a}$

End P Loop

Global summation of ${}^\sigma M_{ia}^P$ over all processes

Form $\{i_{n_p}\}$ by distributing o occupied orbitals over processes

Loop over $i \in \{i_{n_p}\}$

Transform Auxiliary index ${}^\sigma B_{ia}^P = \sum_P {}^\sigma M_{ia}^P (P|Q)^{-\frac{1}{2}}$

End i Loop

Global summation of ${}^\sigma B_{ia}^P$ over all processes

End σ Loop

Form $\{ij_{n_p}\}$ by distributing ij occupied orbitals pairs over processes

Do For $(\sigma_1 = \sigma_2 = \alpha)$; $(\sigma_1 = \sigma_2 = \beta)$; $(\sigma_1 = \alpha, \sigma_2 = \beta)$

Loop over $ij \in \{ij_{n_p}\}$

Calculate $(ia|jb)_{\mathrm{RI}} = \sum_P {}^{\sigma_1} B_{ia}^P \; {}^{\sigma_2} B_{jb}^P$ and load non-RI $(ia|jb)$, (all virtual a, b)

According to (σ_1, σ_2) form t_{ij}^{ab} either using $(ia|jb)_{\mathrm{RI}}$ and $(ia|jb)$

Accumulate $\tilde{E}$ and $E_{\mathrm{RI}}^{(2)}$

End ij Loop

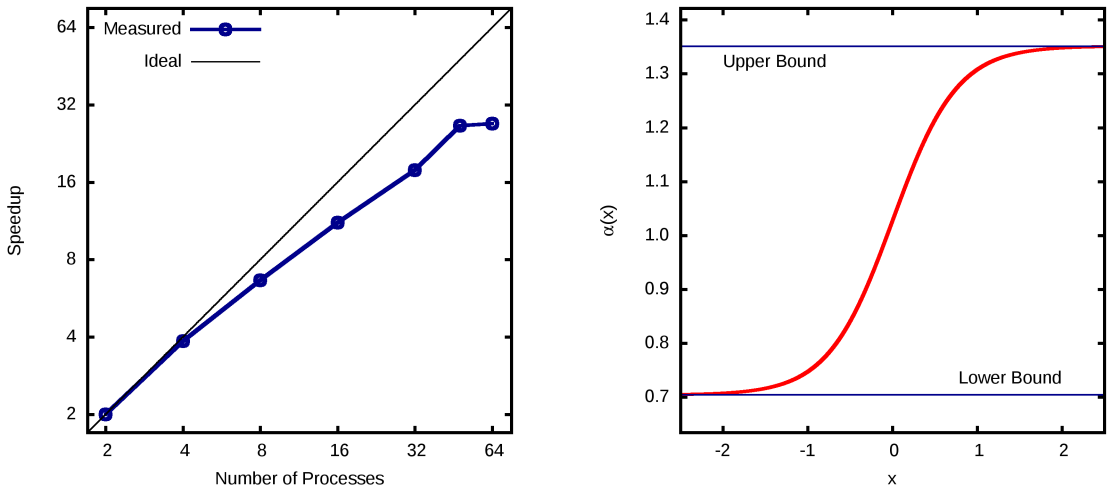
End Do For

Global summation of $\tilde{E}$ and $E_{\mathrm{RI}}^{(2)}$ over all processes

Form $\Delta I = 2\tilde{E} - E_{\mathrm{RI}}^{(2)} - E^{(2)}$ and $|\Delta_{\mathrm{RI}}| = |E^{(2)} - E_{\mathrm{RI}}^{(2)}|$

Figure B.1: Pseudocode of the parallel algorithm for computing ΔI.

In order to ensure convergence of the optimization problem it is necessary to provide an initial guess that fulfill certain requirements. In particular, since only non-contracted basis functions are considered, the parameters that must be defined at first are the maximum



(a) **(b)**

Figure B.2: (a) Measured speed up with respect to 2 processes for performing 20 steps of RI basis optimization for the Ti atom (55 primary basis function, 165 auxiliary basis function, 33 exponents optimized, $N_w = 2$). (b) Plot of the transformation function $\alpha(x)$ for the exponent of the single i function in the auxiliary RI basis of the Ti atom.

angular momentum and the number of shell for each of them. Moreover the number of steps needed to convergence can be significantly reduced if the initial values for the $\{\alpha_i\}$ are properly chosen. It is clear that if all possible pair product of Gaussian functions of the primary basis are included, then the auxiliary basis is complete and there is not need for any optimization. Unfortunately this choice will not bring any computational savings from the RI approximation. Therefor, as a general rule, the size of the RI basis should not be larger than four times the size of the primary basis. Nevertheless the analysis of the pair product of the primary basis functions can be used to extract information for the generation of a good initial guess.

In the present implementation three possibilities have been considered:

1. The initial RI basis is provided in input.

2. The number of shell for each quantum number up to the maximum are defined in input, but not the values of the exponents.

3. The initial RI basis is automatically generated.

The first case is offering the maximum flexibility, but needs a sufficient knowledge of the problem as well as enough experience in basis set optimization. In the second case, given the n_l exponents for the l quantum number, the $\{\alpha_i(l)\}$ are defined by the following progression:

$$\alpha_i(l) = \begin{cases} \alpha_{\min}(l) \cdot g^{i-1}, \quad g = \left(\frac{\alpha_{\max}(l)}{\alpha_{\min}(l)}\right)^{1/(n_l-1)} & \text{if} \quad n_l > 1 \\ \left(\alpha_{\max}(l) - \alpha_{\min}(l)\right)/2 & \text{if} \quad n_l = 1 \end{cases} \tag{B.12}$$

Table B.1: Reported are the basis parameters, exponents (α_i) and contraction coefficients (c_i), at the triple-zeta level (cc-TZ) for the Titanium atom. The electronic wavefunction has been converged with the UHF method based on the pseudopotential of the GTH form [101]. The number of valence electrons treated is 12, the considered spin multiplicity is 3. All electrons have been correlated in the MP2 energy calculation. For the RI auxiliary basis the quantities used for monitoring the convergence are reported at the bottom of the table for the automatically generated initial guess and for the converged basis. In this case the contraction coefficients are omitted since the basis is made of only primitive Gaussian's (implicitly $c_i = 1.0$).

Primary cc-TZ Basis				RI-TZ Basis		
Shell Type	α_i	c_i		Shell Type	α_i Initial Guess	α_i Optimized
s	3.798 670	−0.105 315	−0.682 730	s	0.324 000	0.219 130
	1.911 401	0.212 651	0.694 805	s	0.496 720	0.346 878
	0.750 239	−0.971 436	0.226 111	s	0.759 171	0.558 168
sp	0.300 000	1.000 000		s	1.160 293	0.907 894
sp	0.200 000	1.000 000		s	1.773 356	1.470 988
spd	0.130 000	1.000 000		s	2.710 341	2.360 246
p	9.460 777	−0.099 229		s	4.142 401	3.749 257
	1.819 086	0.670 782		s	6.331 116	5.907 634
	0.712 525	0.734 986		p	0.324 993	0.191 489
d	5.836 423	0.299 406		p	0.584 953	0.280 561
	2.276 618	0.560 599		p	1.052 851	0.536 046
	0.904 942	0.772 065		p	1.895 011	1.095 540
d	0.363 258	1.000 000		p	3.410 799	2.435 505
f	1.248 300	1.000 000		p	6.139 035	5.274 956
f	0.283 600	1.000 000		p	11.049 531	10.262 333
g	0.725 100	1.000 000		d	0.324 991	0.200 263
				d	0.706 408	0.501 345
				d	1.535 455	0.957 144
				d	3.337 468	1.734 426
				d	7.254 317	4.170 611
				d	15.767 954	10.948 096
				f	0.325 002	0.350 815
				f	0.813 334	0.748 129
				f	2.035 436	1.500 011
				f	5.093 830	3.455 743
				f	12.747 667	9.526 201
				g	0.325 000	0.450 106
				g	1.009 068	1.001 964
				g	3.132 952	2.124 931
				g	9.727 365	5.638 426
				h	0.517 155	0.886 458
				h	8.487 172	1.887 485
				i	3.088 470	1.292 600
$E^{(2)} = -0.335\,604$				$E^{(2)}_{\mathrm{RI}} =$	$-0.335\,476$	$-0.335\,600$
				$\lvert\Delta_{\mathrm{RI}}\rvert =$	$1.28 \cdot 10^{-4}$	$4.16 \cdot 10^{-6}$
				$\Delta I =$	$3.12 \cdot 10^{-4}$	$2.99 \cdot 10^{-7}$
				$\Delta I^{\mathrm{REL}} =$	$9.29 \cdot 10^{-4}$	$8.90 \cdot 10^{-7}$

where $\alpha_{\max}(l)$ and $\alpha_{\min}(l)$ are scaled maximum and minimum exponents extracted by inspection of the pair product of the primary basis functions with resulting angular momentum equal to l. For the last case, it is difficult to define general rules to systematically

generate an initial RI basis given an arbitrary primary basis. Three possibilities can be chosen for the generation, namely "MEDIUM", "LARGE" and "VERY_LARGE", controlling the size of the initial guess. For this purpose the number of shells for each quantum number is defined with simple empirical rules chosen according to the features of the primary basis. The initial values of the exponents are set using the same progression formula previously described (Equation (B.12)). It is clear that the usage of the automatically generated initial guess can be useful for a preliminary study of the problem, but it doesn't guarantee that this choice will provide convergence in the optimization process. In Table B.1 the automatically generated RI basis as well as the final converged result are reported together with the values of the quantities used for monitoring the convergence. The RI basis is generated for a primary basis at the triple-zeta level for Ti (also reported in the Table B.1).

One of the issue that was originally encountered is related to the fact that, if the initial values of the $\{\alpha_i\}$ were not properly chosen, then during the optimization procedure, two or more exponents were getting too close leading to linear dependence in the RI basis, with consequent failure in the evaluation of $(P|Q)^{-1/2}$. This may be due to the low accuracy in the numerical calculation of the derivatives. Moreover, the numerical evaluation of the derivatives introduces complications in the definition of the $\delta\alpha_i$ (step size) when the exponents are very different among each others. In order to overcome these problems, instead of directly dealing with the bare exponents, the optimization is carried out with respect to a new variable x according to the following mapping:

$$\alpha_i(x) : \mathbb{R} \to [L_i : U_i] \tag{B.13}$$

$$\alpha_i(x) = L_i + \frac{U_i - L_i}{1 + e^{-a \cdot x}} \tag{B.14}$$

where L_i and U_i are respectively the lower and upper bound for the exponent α_i and a is an optimal value chosen such that the curvature of $\alpha_i(x)$ is an extremum point for $x = 0.5$ and $x = -0.5$. The transformation function for the exponent of the single i function in the auxiliary RI basis of the Ti atom is shown in Figure B.2b. The values of L_i and U_i are chosen such that two consecutive exponents for each angular momentum can never overlap. Additionally, during the optimization procedure, an internal check for each angular momentum is performed to assess if the range of definition of the exponents is still adequate for the actual stage of the optimization. If this is not the case, all ranges for the given angular momentum are redefined shrinking or enlarging them according to the values of the exponents.

Finally, it has to be taken into account that, even for a good initial guess, the optimization can take $O(100)$ steps. Possible reasons for lack of convergence can be due to:

- Too few or too many auxiliary functions.
- The RI basis that is not balanced with respect to the primary, *i.e.* too many exponents of one angular momentum with respect to the others.
- Numerical evaluation of the derivatives not accurate enough. In this case the step size can be decreased employing the associated key in the input.
- The maximum angular momentum of the RI basis is too low.

Appendix C

Ice XV: Computational Setups, Structures, and Validation

Computational Details

CASTEP settings

The CASTEP [339] reference PBE calculations were undertaken using version 7.0, a cutoff of 800eV, on-the-fly pesudopotentials and k-point sampling at 0.03 Angstroms^{-1} (108 k-points). Under an isotropic external pressure of 1GPa, each of the 18 unit cells, containing 10 molecules, was fully relaxed until mechanical equilibrium was attained. The lattice parameters a,b,c and α, β, γ were treated as independent variables in the cell optimization using the BFGS scheme.

CP2K settings

The calculations performed with the CP2K program [90] are all based on the Gaussian and Plane-Waves (GPW) method [86, 91] that makes use of a Gaussian basis to expand molecular orbitals and an auxiliary plane wave basis for the expansion of the electronic density. In the GPW scheme core electrons are replaced by pseudopotentials in order for the electronic density to be efficiently expanded in the plane waves basis. The employed pseudopotentials are of dual-space, Goedecker-Teter-Hutter (GTH) type [101]. In particular, pseudopotentials have been specifically generated for the Hartree-Fock (HF), PBE and PBE0 methods, for all other DFT calculations the PBE and PBE0 pseudopotentials have been used respectively for methods of GGA or Hybrid type. Pseudopotential parameters can be retrieved from supporting information of Ref. [177].

The non-local HF exchange energy is calculated by employing a Γ-point method based on the usage of a truncated Coulomb operator to avoid divergences in the energy when periodic boundary conditions (PBC) are considered [104, 105]. For wide bandgap systems, this allows for stable calculations in the condensed phase, without loss of accuracy, if a truncation radius larger than $5 - 6$ Å is used [105, 106]. The employed truncation radius is 5.5 Å.

The calculation of the MP2 and RPA energies within the RI-GPW [131] approach (RI-MP2, RI-RPA) is closely related to the original GPW method. In this case, the dual

representation of the electronic density is applied to the fitting density arising from the resolution of identity approximation (RI) within the Coulomb metric [69, 144–146]. The RI method requires the introduction of an auxiliary Gaussian basis, the RI basis have been specifically generated tailored on the employed primary basis. In the RI-RPA method, the correlation energy is given in term of a frequency integral, the integration is carried out employing the Clenshaw-Curtius quadrature scheme using 40 grid points.

The employed primary basis have been labeled as cc-TZ, cc-QZ and cc-5Z denoting respectively valence-only correlation-consistent type basis [102, 103] of triple, quadruple and quintuple zeta level, generated for being used with pseudopotentials. These basis contain functions with angular momentum up to f, g and h for Oxygen and d, f and g for Hydrogen respectively for the cc-TZ, cc-QZ and cc-5Z basis. The associated RI auxiliary basis (RI-TZ, RI-QZ and RI-5Z) has been generated following the procedure proposed by Weigend *et al.* [69]. These primary and auxiliary basis sets can be retrieved from supporting information of Ref. [308].

The number of grid points used in the GPW scheme for the representation of the electronic density is kept constant for all calculations. This avoids discontinuities in the energy due to the volume change in the cell relaxations. Additionally, this is a safe choice when comparing the energy computed on structures with different volume. The original grids are constructed for the reference cell with density $1.089\,\mathrm{g/cm^3}$ corresponding to a $2\times2\times2$ supercell with edge length of 13.00 Å. The wavefunction optimization is performed employing a charge-density cutoffs (E_{cut}) of 800, 1200 and 1600 Ry, respectively for the HF, GGA-DFT and meta-GGA-DFT methods. E_{cut} is progressively increased in order to ensure convergence in the calculation of the Hartree potential, gradients of the density and the kinetic energy density, respectively for the HF, GGA-DFT and meta-GGA-DFT methods at low cost compared to the evaluation of the HF exchange or non-local dynamical correlation energies. For the RI-MP2 and RI-RPA calculations a cutoff of 300 Ry is used while for the non-local part of the van der Waals density functionals $E_{\mathrm{cut}}^{\mathrm{NL}}$ is set to 600 Ry.

For all level of theory considered, each of the 18 cells have been fully relaxed (a,b,c and α, β, γ were treated as independent variables) under 1 GPa isotropic external pressure using the BFGS scheme until mechanical equilibrium was established employing the cc-TZVP basis [44, 131, 177].

Basis Set Analysis

For all calculations at the DFT level, unless otherwise specified, the cc-TZ basis has been used, that provides converged results at this level of theory. In fact, as shown in Figure C.1 for the PBE case, the cc-TZ provides relative stability over the 18 structures of ice XV that are basically identical as those obtained with the cc-QZ basis. In this respect the cc-TZ basis can be considered of similar quality as the CP2K standard QZV3P basis, that has been shown to provide very well converged results at the DFT level [203, 294] for liquid water in the condensed phase.

Different is the case for the calculation of the RI-MP2 and RI-RPA correlation energies. In this case, due to the electron coalescence cusp [120, 133, 134], the convergence of the correlation energies with respect to the basis set is significantly slower compared the the DFT cases. This problem has been investigated by analyzing the convergence of the

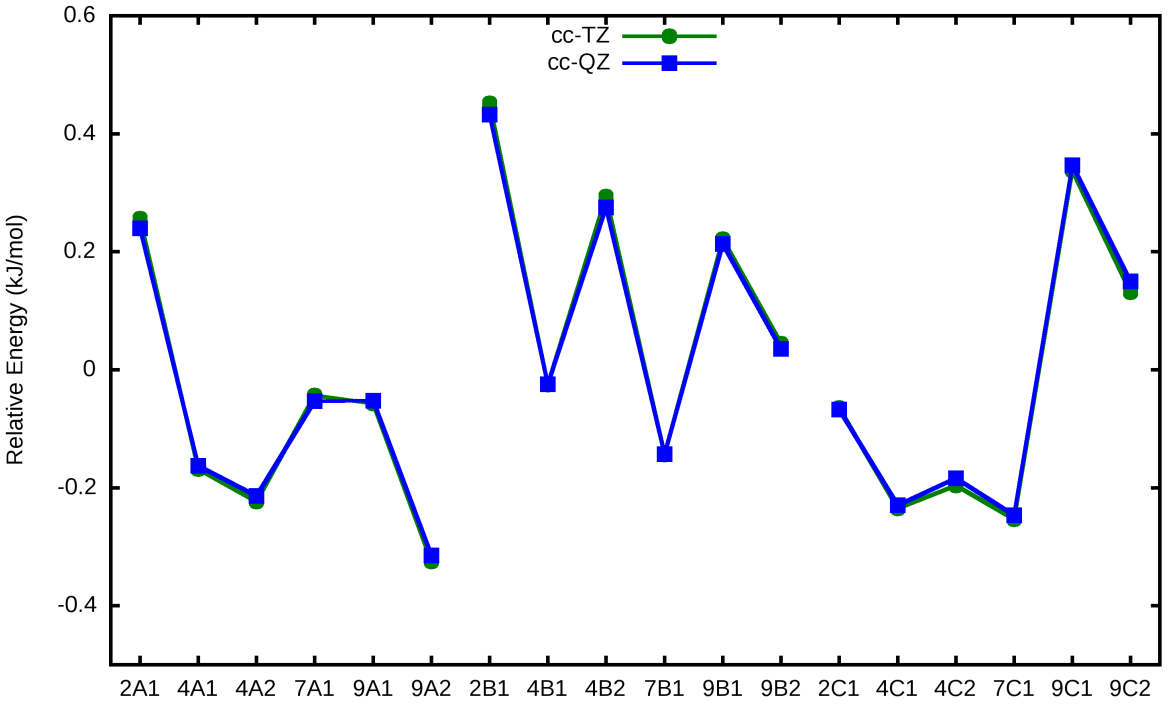


Figure C.1: Relative energies, with respect to the average, of the 18 symmetry inequivalent structures possible in the ice XV unit cell reported using PBE, cc-TZ and cc-QZ basis.

Table C.1: Representative timings for the evaluation of the correlation energies (2C1 structure) with the various basis employed in this work. The system is made of 80 H_2O molecules ($2 \times 2 \times 2$ supercell) resulting in 320 occupied molecular orbitals. Timing measured on a CRAY-XC30 machine, each hybrid node mounts one graphical processing unit (NVIDIA-Tesla-K20X) and 8 processors (Intel-Xeon-E5).

	cc-TZ	cc-QZ	cc-5Z
Number of Basis Functions	4560	9120	16000
Number of RI Basis Functions	10880	19040	29600
Number of Hybrid Nodes	1024	2048	4096
Time for RI-RPA Energy (s)	216.9	347.6	1165.4
Time for RI-MP2 Energy (s)	279.1	523.8	-

relative energies of the ice XV structures with respect of the basis set quality over the series cc-TZ, cc-QZ and cc-5Z. This analysis has been focused on the RPA correlation energy that shows essentially the same basis set convergence behavior as the MP2 energy [43, 133]. This choice is motivated by the favorable scaling of RPA compared to MP2, that is $O(N^4)$ *vs* $O(N^5)$.

In general, for correlation-consistent type basis the correlation energies show an inverse cubic dependence $E^c \propto X^{-3}$ with respect to the basis quality cc-XZ [120]. The results for the relative stability of the 18 structures of ice XV as obtained using the cc-TZ, cc-QZ

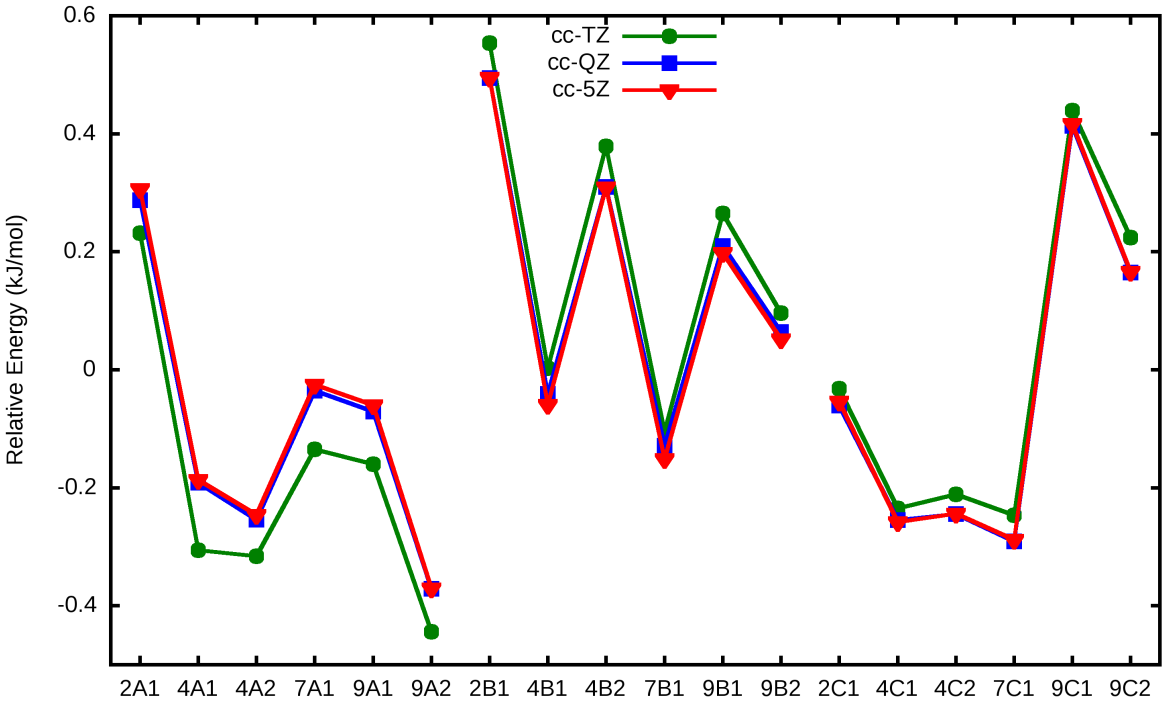


Figure C.2: Relative energies, with respect to the average, of the 18 symmetry inequivalent structures possible in the ice XV unit cell reported using the cc-TZ, cc-QZ and cc-5Z basis at the RPA level.

and cc-5Z basis at the RPA level are shown in Figure C.2. From the plot it appears that the usage of the cc-TZ basis provides results qualitatively in agreement with the higher quality basis while the cc-QZ and cc-5Z give basically the same profiles. These results shows that for the relative stability of ice XV structures the usage of the cc-QZ basis provides essentially converged results for the correlation energies evaluation.

Representative timings for the correlation energies evaluations are reported in Table C.1 for the different calculations reported here.

Supplementary Calculations

The results in Figure C.3 shows that various selected VdWDF's yield the same qualitative order of energies for the ice XV structures as PBE based approaches. The tested functionals are vdw-DF [360, 361], vdw-DF2 [340], optPBE-B88vdw [341, 362] and rVV10 [342].

In Figure C.4 it is shown that despite substantial variation in the proportion of HF exchange used between the different formulations of M06 functionals [343], these methods consistently yield the same overall structure/energy relationship. In the plot, the D3 label stands for Grimme's van der Waals correction [124].

The dipole moment of the different cells is reported in Figure C.5 showing that only

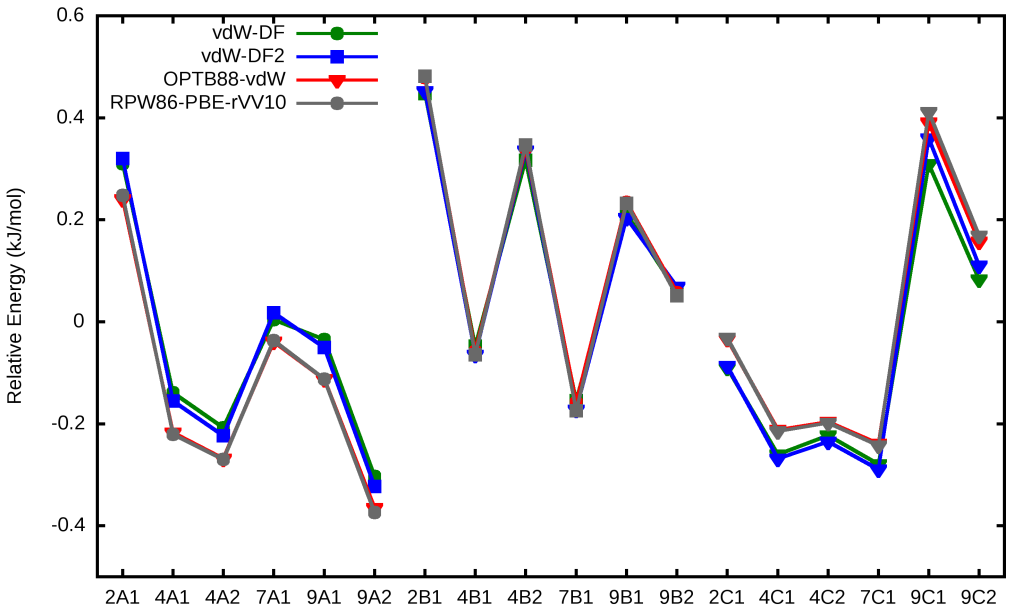


Figure C.3: Relative energies, with respect to the average, of the 18 symmetry inequivalent structures possible in the ice XV unit cell reported using van der Waals density based functionals.

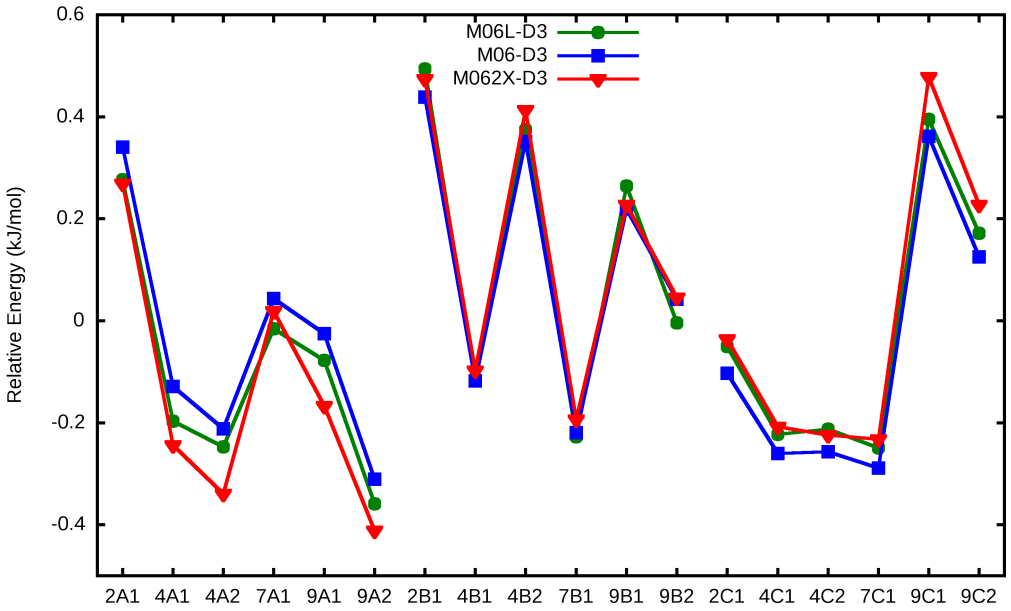


Figure C.4: Relative energies, with respect to the average, of the 18 symmetry inequivalent structures possible in the ice XV unit cell reported using M06 approaches.

four structures are anti-ferroelectric (apolar): 2A1, 2B1, 2C1 and 9A1. More precisely, 2A1, 2B1 and 2C1 have exactly 0 dipole while 9A1 and 9B2 have very small residual dipoles.

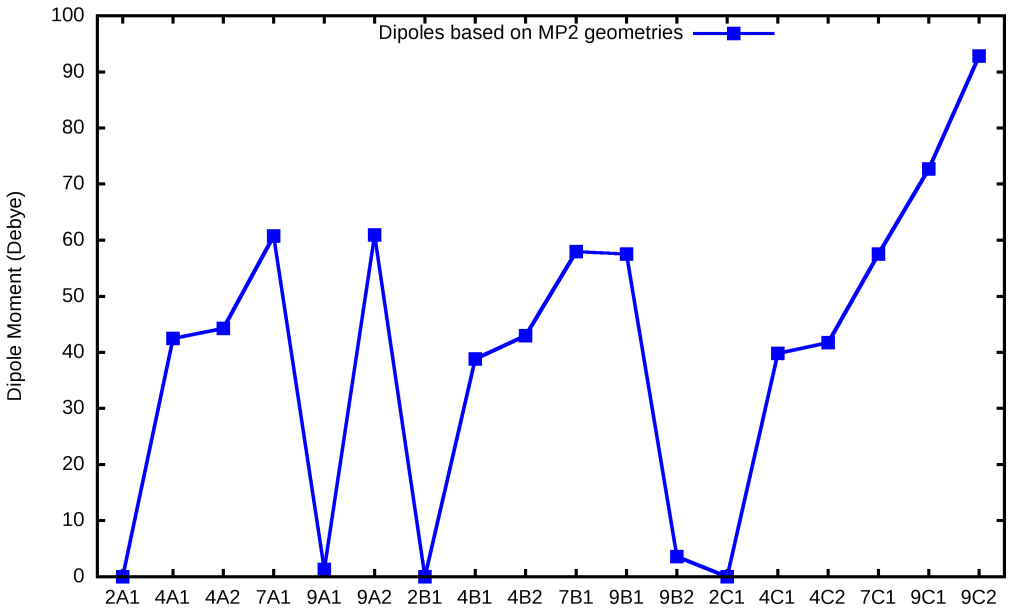


Figure C.5: The dipole moments of the 18 symmetry inequivalent structures possible in the ice XV unit cell.

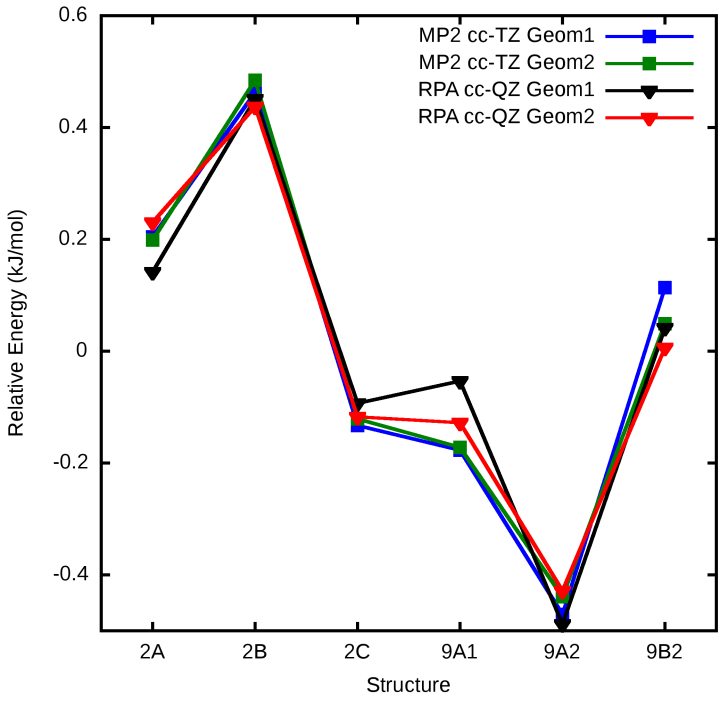


Figure C.6: Optimized energies of five XV structures at the MP2 and RPA level calculated using the geometries from Ref. [309] (Geom1) and those obtained here (Geom2).

To assess whether the initial geometries influence the optimized structures and energy differences, full cell and coordinate optimizations have been performed using the MP2 geometries obtained in this work and those reported in the supporting information of Ref. [309]. Figure C.6 shows a very minor influence of the starting geometry on the final RPA and MP2 relative energies obtained. The MP2 and RPA absolute energies obtained from MP2 geometries obtained here are found to have lower absolute energy than those obtained using structures from Ref. [309].

Ice XV Inequivalent Structures

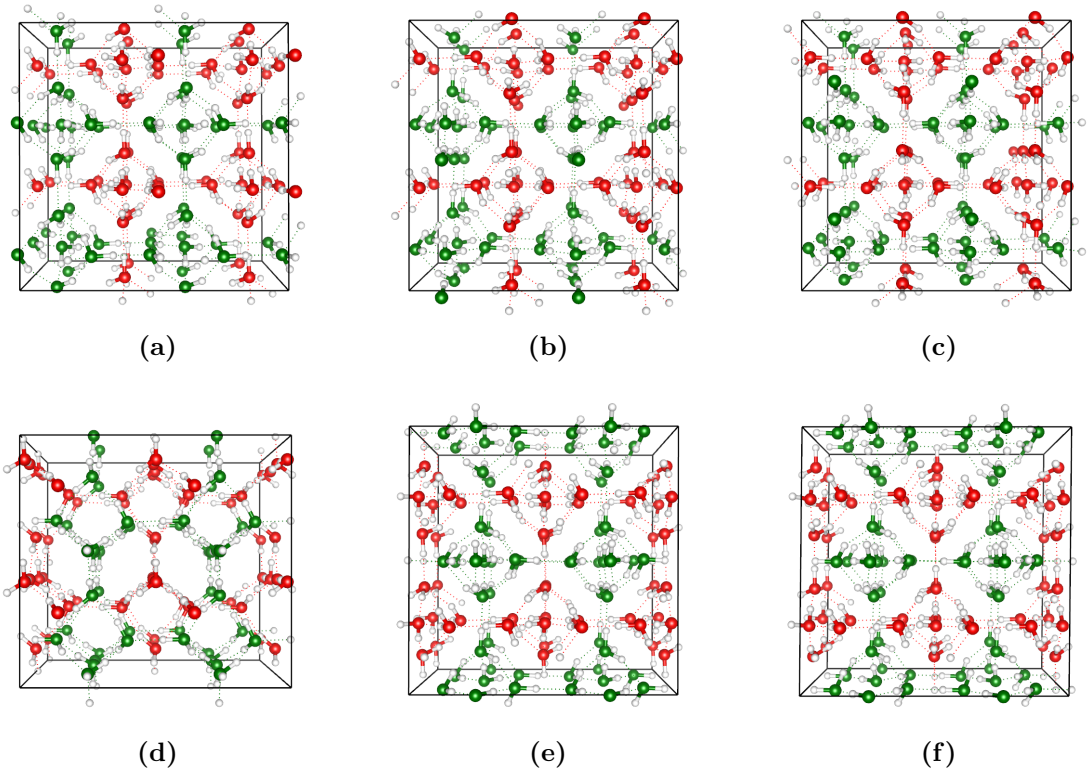


Figure C.7: A structures: (a) 2A1, (b) 4A1, (c) 4A2, (d) 7A1, (e) 9A1, (f) 9A2.

Reported here are the optimized geometries at the MP2 level for the 18 symmetry inequivalent structures possible in the ice XV. In all cases, Hydrogen is shown white spheres while Oxygen atoms are shown in green and red in order to distinguish the two sub-lattices within the structure. Each figure shows a $2 \times 2 \times 2$ supercell along the [001] axis. The inequivalent structures are grouped using the same criterion as that employed in reporting the relative energies, that is Figure C.7, C.8 and C.9 show respectively the structures labeled with A, B and C.

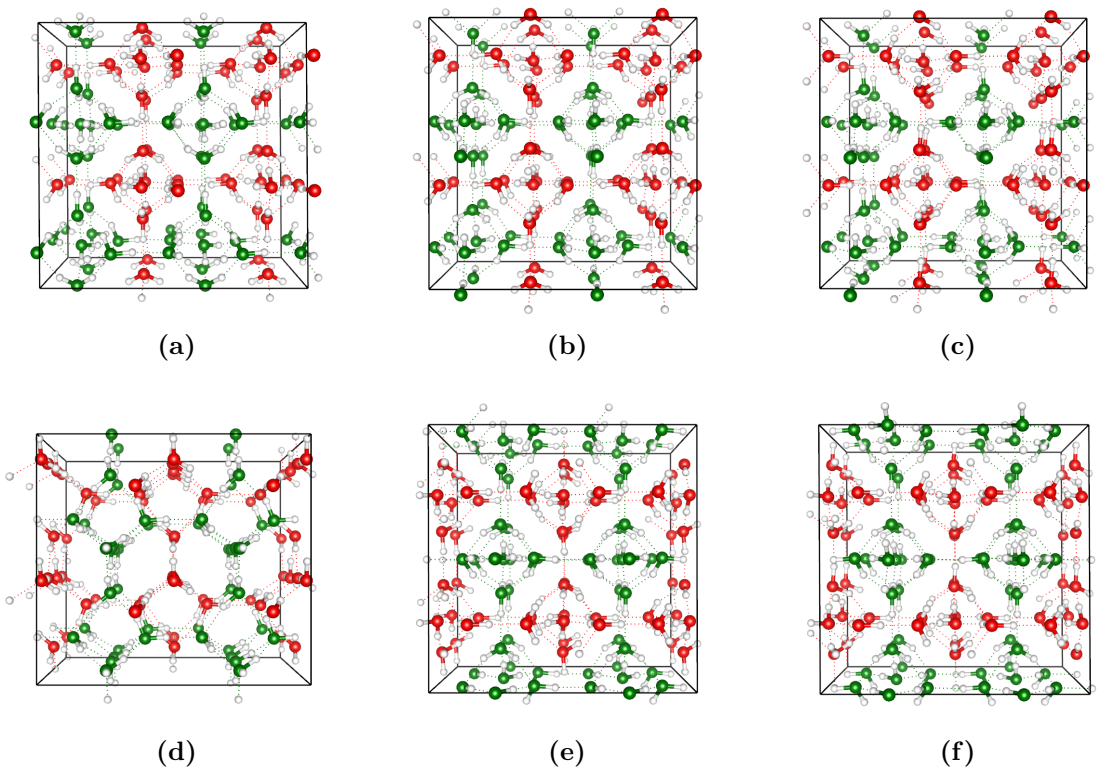


Figure C.8: B structures: (a) 2B1, (b) 4B1, (c) 4B2, (d) 7B1, (e) 9B1, (f) 9B2.

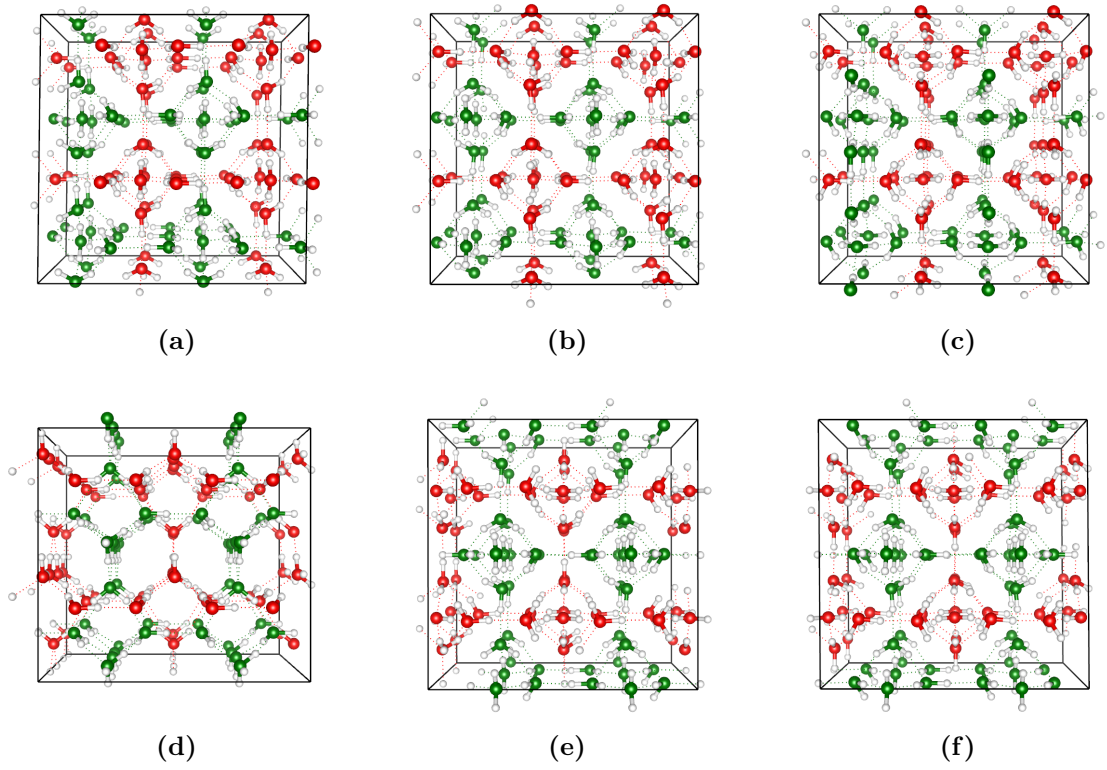


Figure C.9: C structures: (a) 2C1, (b) 4C1, (c) 4C2, (d) 7C1, (e) 9C1, (f) 9C2.

Bibliography

[1] P. Hohenberg and W. Kohn "Inhomogeneous Electron Gas" *Phys. Rev.* **136**, B864 (1964).

[2] W. Kohn and L. J. Sham "Self-Consistent Equations Including Exchange and Correlation Effects" *Phys. Rev.* **140**, A1133 (1965).

[3] J. P. Perdew, A. Ruzsinszky, J. Tao, V. N. Staroverov, G. E. Scuseria, and G. I. Csonka "Prescription for the design and selection of density functional approximations: More constraint satisfaction with fewer fits" *J. Chem. Phys.* **123**, 062201 (2005).

[4] D. C. Langreth and J. P. Perdew "Exchange-correlation energy of a metallic surface: Wave-vector analysis" *Phys. Rev. B* **15**, 2884 (1977).

[5] D. Langreth and J. Perdew "The exchange-correlation energy of a metallic surface" *Solid State Commun.* **17**, 1425 (1975).

[6] O. Gunnarsson and B. I. Lundqvist "Exchange and correlation in atoms, molecules, and solids by the spin-density-functional formalism" *Phys. Rev. B* **13**, 4274 (1976).

[7] M. Fuchs and X. Gonze "Accurate density functionals: Approaches using the adiabatic-connection fluctuation-dissipation theorem" *Phys. Rev. B* **65**, 235109 (2002).

[8] F. Furche and T. V. Voorhis "Fluctuation-dissipation theorem density-functional theory" *J. Chem. Phys.* **122**, 164106 (2005).

[9] A. Heßelmann and A. Görling "Random-phase approximation correlation methods for molecules and solids" *Mol. Phys.* **109**, 2473 (2011).

[10] H. Eshuis, J. Bates, and F. Furche "Electron correlation methods based on the random phase approximation" *Theor. Chem. Acc.* **131**, 1084 (2012).

[11] J. Paier, X. Ren, P. Rinke, G. E. Scuseria, A. Grüneis, G. Kresse, and M. Scheffler "Assessment of correlation energies based on the random-phase approximation" *New J. Phys.* **14**, 043002 (2012).

[12] X. Ren, P. Rinke, C. Joas, and M. Scheffler "Random-phase approximation and its applications in computational chemistry and materials science" *J. Mater. Sci.* **47**, 7447 (2012).

[13] C. Møller and M. S. Plesset "Note on an Approximation Treatment for Many-Electron Systems" *Phys. Rev.* **46**, 618 (1934).

[14] A. Szabo and N. S. Ostlund *Modern Quantum Chemistry* McGraw Hill, New York (1982).

[15] A. Görling and M. Levy "Correlation-energy functional and its high-density limit obtained from a coupling-constant perturbation expansion" *Phys. Rev. B* **47**, 13105 (1993).

[16] R. J. Bartlett, I. Grabowski, S. Hirata, and S. Ivanov "The exchange-correlation potential in ab initio density functional theory" *J. Chem. Phys.* **122**, 034104 (2005).

[17] S. Grimme "Semiempirical hybrid density functional with perturbative second-order correlation" *J. Chem. Phys.* **124**, 034108 (2006).

[18] T. Schwabe and S. Grimme "Towards chemical accuracy for the thermodynamics of large molecules: new hybrid density functionals including non-local correlation effects" *Phys. Chem. Chem. Phys.* **8**, 4398 (2006).

[19] L. Goerigk and S. Grimme "Efficient and Accurate Double-Hybrid-Meta-GGA Density Functional-sEvaluation with the Extended GMTKN30 Database for General Main Group Thermochemistry, Kinetics, and Noncovalent Interactions" *J. Chem. Theory Comput.* **7**, 291 (2011).

[20] T. Helgaker, P. Jørgensen, and J. Olsen *Molecular Electronic Structure Theory* John Wiley & Sons, Chichester (2000).

[21] R. G. Parr and W. Yang *Density-Functional Theory of Atoms and Molecules* Oxford Science Publications, New York (1989).

[22] W. Koch and M. C. Holthausen *A Chemist's Guide to Density Functional Theory* Wiley-VCH Verlag GmbH (2001).

[23] F. Jensen *Introduction to Computational Chemistry* John Wiley & Sons, Chichester (2007).

[24] M. Born and R. Oppenheimer "Zur Quantentheorie der Molekeln" *Annalen der Physik* **84**, 0457 (1927).

[25] J. C. Slater "The Theory of Complex Spectra" *Phys. Rev.* **34**, 1293 (1929).

[26] E. U. Condon "The Theory of Complex Spectra" *Phys. Rev.* **36**, 1121 (1930).

[27] C. C. J. Roothaan "New Developments in Molecular Orbital Theory" *Rev. Mod. Phys.* **23**, 69 (1951).

[28] K. Raghavachari and J. B. Anderson "Electron Correlation Effects in Molecules" *J. Phys. Chem.* **100**, 12960 (1996).

[29] J. A. Pople, J. S. Binkley, and R. Seeger "Theoretical models incorporating electron correlation" *Int. J. Quantum Chem. Quantum Chem. Symp.* **10**, 1 (1976).

[30] S. Hirata, X. He, M. R. Hermes, and S. Y. Willow "Second-Order Many-Body Perturbation Theory: An Eternal Frontier" *J. Phys. Chem. A* **118**, 655 (2014).

[31] R. A. Harris "Oscillator Strengths and Rotational Strengths in HartreeFock Theory" *J. Chem. Phys.* **50**, 3947 (1969).

[32] P. Jørgensen "Molecular and Atomic Applications of Time-Dependent Hartree-Fock Theory" *Ann. Rev. Phys. Chem.* **26**, 359 (1975).

[33] A. D. McLachlan and M. A. Ball "Time-Dependent Hartree-Fock Theory for Molecules" *Rev. Mod. Phys.* **36**, 844 (1964).

[34] G. E. Scuseria, T. M. Henderson, and D. C. Sorensen "The ground state correlation energy of the random phase approximation from a ring coupled cluster doubles approach" *J. Chem. Phys.* **129**, 231101 (2008).

[35] F. Furche "Developing the random phase approximation into a practical post-Kohn–Sham correlation model" *J. Chem. Phys.* **129**, 114105 (2008).

[36] W. Klopper, A. M. Teale, S. Coriani, T. B. Pedersen, and T. Helgaker "Spin flipping in ring-coupled-cluster-doubles theory" *Chem. Phys. Lett.* **510**, 147 (2011).

[37] A. D. Becke "Perspective: Fifty years of density-functional theory in chemical physics" *J. Chem. Phys.* **140**, 18A301 (2014).

[38] A. D. Becke "A multicenter numerical integration scheme for polyatomic molecules" *J. Chem. Phys.* **88**, 2547 (1988).

[39] O. Treutler and R. Ahlrichs "Efficient molecular numerical integration schemes" *J. Chem. Phys.* **102**, 346 (1995).

[40] C. W. Murray, N. C. Handy, and G. J. Laming "Quadrature schemes for integrals of density functional theory" *Mol. Phys.* **78**, 997 (1993).

[41] F. Furche "On the density matrix based approach to time-dependent density functional response theory" *J. Chem. Phys.* **114**, 5982 (2001).

[42] F. Furche "Molecular tests of the random phase approximation to the exchange-correlation energy functional" *Phys. Rev. B* **64**, 195120 (2001).

[43] H. Eshuis, J. Yarkony, and F. Furche "Fast computation of molecular random phase approximation correlation energies using resolution of the identity and imaginary frequency integration" *J. Chem. Phys.* **132**, 234114 (2010).

[44] M. Del Ben, J. Hutter, and J. VandeVondele "Second-Order Møller–Plesset Perturbation Theory in the Condensed Phase: An Efficient and Massively Parallel Gaussian and Plane Waves Approach" *J. Chem. Theory Comput.* **8**, 4177 (2012).

[45] J.-Q. Sun and R. J. Bartlett "Second-order many-body perturbation-theory calculations in extended systems" *J. Chem. Phys.* **104**, 8553 (1996).

[46] M. Katouda and S. Nagase "Application of second-order Møller–Plesset perturbation theory with

resolution-of-identity approximation to periodic systems" *J. Chem. Phys.* **133**, 184103 (2010).

[47] A. F. Izmaylov and G. E. Scuseria "Resolution of the identity atomic orbital Laplace transformed second order Møller Plesset theory for nonconducting periodic systems" *Phys. Chem. Chem. Phys.* **10**, 3421 (2008).

[48] P. Y. Ayala, K. N. Kudin, and G. E. Scuseria "Atomic orbital Laplace-transformed second-order Møller–Plesset theory for periodic systems" *J. Chem. Phys.* **115**, 9698 (2001).

[49] C. Pisani, M. Busso, G. Capecchi, S. Casassa, R. Dovesi, L. Maschio, C. Zicovich-Wilson, and M. Schütz "Local-MP2 electron correlation method for nonconducting crystals" *J. Chem. Phys.* **122**, 094113 (2005).

[50] M. Marsman, A. Grüneis, J. Paier, and G. Kresse "Second-order Møller–Plesset perturbation theory applied to extended systems. I. Within the projector-augmented-wave formalism using a plane wave basis set" *J. Chem. Phys.* **130**, 184103 (2009).

[51] A. Grüneis, M. Marsman, and G. Kresse "Second-order Møller–Plesset perturbation theory applied to extended systems. II. Structural and energetic properties" *J. Chem. Phys.* **133**, 074107 (2010).

[52] D. Cremer "MøllerPlesset perturbation theory: from small molecule methods to methods for thousands of atoms" *WIREs Comput. Mol. Sci.* **1**, 509 (2011).

[53] S. Saebø and P. Pulay "Local Treatment of Electron Correlation" *Annu. Rev. Phys. Chem.* **44**, 213 (1993).

[54] P. Pulay and S. Saebø "Orbital-invariant formulation and second-order gradient evaluation in Møller-Plesset perturbation theory" *Theor. Chim. Acta* **69**, 357 (1986).

[55] G. Rauhut, P. Pulay, and H.-J. Werner "Integral transformation with low-order scaling for large local second-order MøllerPlesset calculations" *J. Comput. Chem.* **19**, 1241 (1998).

[56] M. Schütz, G. Hetzer, and H.-J. Werner "Low-order scaling local electron correlation methods. I. Linear scaling local MP2" *J. Chem. Phys.* **111**, 5691 (1999).

[57] G. Hetzer, M. Schütz, H. Stoll, and H.-J. Werner "Low-order scaling local correlation methods II: Splitting the Coulomb operator in linear scaling local second-order Møller-Plesset perturbation theory" *J. Chem. Phys.* **113**, 9443 (2000).

[58] S. Saebø and P. Pulay "A low-scaling method for second order Møller-Plesset calculations" *J. Chem. Phys.* **115**, 3975 (2001).

[59] C. Pisani, L. Maschio, S. Casassa, M. Halo, M. Schütz, and D. Usvyat "Periodic local MP2 method for the study of electronic correlation in crystals: Theory and preliminary applications" *J. Comput. Chem.* **29**, 2113 (2008).

[60] P. Maslen "Non-iterative local second order Møller-Plesset theory" *Chem. Phys. Lett.* **283**, 102 (1998).

[61] P. E. Maslen and M. Head-Gordon "Noniterative local second order Møller-Plesset theory: Convergence with local correlation space" *J. Chem. Phys.* **109**, 7093 (1998).

[62] J. Almlöf "Elimination of energy denominators in Møller-Plesset perturbation theory by a Laplace transform approach" *Chem. Phys. Lett.* **181**, 319 (1991).

[63] M. Häser and J. Almlöf "Laplace transform techniques in Møller-Plesset perturbation theory" *J. Chem. Phys.* **96**, 489 (1992).

[64] M. Häser "Møller-Plesset (MP2) perturbation theory for large molecules" *Theor. Chim. Acta* **87**, 147 (1993).

[65] P. Y. Ayala and G. E. Scuseria "Linear scaling second-order Møller-Plesset theory in the atomic orbital basis for large molecular systems" *J. Chem. Phys.* **110**, 3660 (1999).

[66] D. S. Lambrecht, B. Doser, and C. Ochsenfeld "Rigorous integral screening for electron correlation methods" *J. Chem. Phys.* **123**, 184102 (2005).

[67] B. Doser, D. S. Lambrecht, J. Kussmann, and C. Ochsenfeld "Linear-scaling atomic orbital-based second-order Møller–Plesset perturbation theory by rigorous integral screening criteria" *J. Chem. Phys.* **130**, 064107 (2009).

[68] M. Feyereisen, G. Fitzgerald, and A. Komornicki "Use of approximate integrals in ab initio theory. An application in MP2 energy calculations" *Chem. Phys. Lett.* **208**, 359 (1993).

[69] F. Weigend, M. Häser, H. Patzelt, and R. Ahlrichs "RI-MP2: optimized auxiliary basis sets and

demonstration of efficiency" *Chem. Phys. Lett.* **294**, 143 (1998).

[70] D. E. Bernholdt and R. J. Harrison "Fitting basis sets for the RI-MP2 approximate second-order many-body perturbation theory method" *J. Chem. Phys.* **109**, 1593 (1998).

[71] H.-J. Werner, F. R. Manby, and P. J. Knowles "Fast linear scaling second-order Møller-Plesset perturbation theory (MP2) using local and density fitting approximations" *J. Chem. Phys.* **118**, 8149 (2003).

[72] L. Maschio, D. Usvyat, F. R. Manby, S. Casassa, C. Pisani, and M. Schütz "Fast local-MP2 method with density-fitting for crystals. I. Theory and algorithms" *Phys. Rev. B* **76**, 075101 (2007).

[73] D. Usvyat, L. Maschio, F. R. Manby, S. Casassa, M. Schütz, and C. Pisani "Fast local-MP2 method with density-fitting for crystals. II. Test calculations and application to the carbon dioxide crystal" *Phys. Rev. B* **76**, 075102 (2007).

[74] W. Klopper, F. R. Manby, S. Ten-No, and E. F. Valeev "R12 methods in explicitly correlated molecular electronic structure theory" *Int. Rev. Phys. Chem.* **25**, 427 (2006).

[75] A. C. Limaye and S. R. Gadre "A general parallel solution to the integral transformation and second-order Møller-Plesset energy evaluation on distributed memory parallel machines" *J. Chem. Phys.* **100**, 1303 (1994).

[76] A. M. Marquez and M. Dupuis "Parallel computation of the MP2 energy on distributed memory computers" *J. Comput. Chem.* **16**, 395 (1995).

[77] I. M. B. Nielsen and E. T. Seidl "Parallel direct implementations of second-order perturbation theories" *J. Comput. Chem.* **16**, 1301 (1995).

[78] J. Baker and P. Pulay "An efficient parallel algorithm for the calculation of canonical MP2 energies" *J. Comput. Chem.* **23**, 1150 (2002).

[79] K. Ishimura, P. Pulay, and S. Nagase "A new parallel algorithm of MP2 energy calculations" *J. Comput. Chem.* **27**, 407 (2006).

[80] M. Katouda and S. Nagase "Efficient parallel algorithm of second-order Møller-Plesset perturbation theory with resolution-of-identity approximation (RI-MP2)" *Int. J. Quantum Chem.* **109**, 2121 (2009).

[81] B. Doser, D. S. Lambrecht, and C. Ochsenfeld "Tighter multipole-based integral estimates and parallel implementation of linear-scaling AO MP2 theory" *Phys. Chem. Chem. Phys.* **10**, 3335 (2008).

[82] E. F. Valeev and C. L. Janssen "Second-order Møller–Plesset theory with linear R12 terms (MP2-R12) revisited: Auxiliary basis set method and massively parallel implementation" *J. Chem. Phys.* **121**, 1214 (2004).

[83] Y. Nakao and K. Hirao "A local second-order Møller-Plesset method with localized orbitals: A parallelized efficient electron correlation method" *J. Chem. Phys.* **120**, 6375 (2004).

[84] I. M. B. Nielsen and C. L. Janssen "Local Møller–Plesset Perturbation Theory: A Massively Parallel Algorithm" *J. Chem. Theory Comput.* **3**, 71 (2007).

[85] L. Maschio "Local MP2 with Density Fitting for Periodic Systems: A Parallel Implementation" *J. Chem. Theory Comput.* **7**, 2818 (2011).

[86] G. Lippert, J. Hutter, and M. Parrinello "A hybrid Gaussian and plane wave density functional scheme" *Mol. Phys.* **92**, 477 (1997).

[87] T. J. Martinez and E. A. Carter "Pseudospectral Møller–Plesset perturbation theory through third order" *J. Chem. Phys.* **100**, 3631 (1994).

[88] K. Ishimura and S. Ten-no "MPI/OpenMP hybrid parallel implementation of second-order MøllerPlesset perturbation theory using numerical quadratures" *Theor. Chem. Acc.* **130**, 317 (2011).

[89] S. Hirata and S. Iwata "Analytical energy gradients in second-order Møller–Plesset perturbation theory for extended systems" *J. Chem. Phys.* **109**, 4147 (1998).

[90] The CP2K developers group CP2K is freely available from: http://www.cp2k.org/ (2014).

[91] J. VandeVondele, M. Krack, F. Mohamed, M. Parrinello, T. Chassaing, and J. Hutter "Quickstep: Fast and accurate density functional calculations using a mixed Gaussian and plane waves approach" *Comput. Phys. Commun.* **167**, 103 (2005).

[92] G. Aissing and H. J. Monkhorst "On the removal of the exchange singularity in extended systems" *Int. J. Quantum Chem.* **48**, 81 (1993).

[93] G. J. Martyna and M. E. Tuckerman "A reciprocal space based method for treating long range interactions in ab initio and force-field-based calculations in clusters" *J. Chem. Phys.* **110**, 2810 (1999).

[94] P. E. Blöchl "Electrostatic decoupling of periodic images of plane-wave-expanded densities and derived atomic point charges" *J. Chem. Phys.* **103**, 7422 (1995).

[95] L. Genovese, T. Deutsch, A. Neelov, S. Goedecker, and G. Beylkin "Efficient solution of Poisson's equation with free boundary conditions" *J. Chem. Phys.* **125**, 074105 (2006).

[96] G. Lippert, J. Hutter, and M. Parrinello "The Gaussian and augmented-plane-wave density functional method for ab initio molecular dynamics simulations" *Theor. Chem. Acc.* **103**, 124 (1999).

[97] M. Krack and M. Parrinello "All-electron ab-initio molecular dynamics" *Phys. Chem. Chem. Phys.* **2**, 2105 (2000).

[98] J. VandeVondele, U. Borštnik, and J. Hutter "Linear Scaling Self-Consistent Field Calculations with Millions of Atoms in the Condensed Phase" *J. Chem. Theory Comput.* **8**, 3565 (2012).

[99] U. Borštnik, J. VandeVondele, V. Weber, and J. Hutter "Sparse matrix multiplication: The distributed block-compressed sparse row library" *Parallel Comput.* **40**, 47 (2014).

[100] "MPI, Message Passing Interface" `http://www.mpi-forum.org/`.

[101] S. Goedecker, M. Teter, and J. Hutter "Separable dual-space Gaussian pseudopotentials" *Phys. Rev. B* **54**, 1703 (1996).

[102] T. H. Dunning "Gaussian basis sets for use in correlated molecular calculations. I. The atoms boron through neon and hydrogen" *J. Chem. Phys.* **90**, 1007 (1989).

[103] D. E. Woon and T. H. Dunning "Gaussian basis sets for use in correlated molecular calculations. III. The atoms aluminum through argon" *J. Chem. Phys.* **98**, 1358 (1993).

[104] M. Guidon, F. Schiffmann, J. Hutter, and J. VandeVondele "Ab initio molecular dynamics using hybrid density functionals" *J. Chem. Phys.* **128**, 214104 (2008).

[105] M. Guidon, J. Hutter, and J. VandeVondele "Robust Periodic Hartree–Fock Exchange for Large-Scale Simulations Using Gaussian Basis Sets" *J. Chem. Theory Comput.* **5**, 3010 (2009).

[106] J. Paier, C. V. Diaconu, G. E. Scuseria, M. Guidon, J. VandeVondele, and J. Hutter "Accurate Hartree-Fock energy of extended systems using large Gaussian basis sets" *Phys. Rev. B* **80**, 174114 (2009).

[107] L. Maschio, D. Usvyat, M. Schütz, and B. Civalleri "Periodic local Møller–Plesset second order perturbation theory method applied to molecular crystals: Study of solid NH_3 and CO_2 using extended basis sets" *J. Chem. Phys.* **132**, 134706 (2010).

[108] F. H. Allen "The Cambridge Structural Database: a quarter of a million crystal structures and rising." *Acta Crystallogr., Sect. B: Struct. Sci* **58**, 380 (2002).

[109] A. D. Becke "Density-functional exchange-energy approximation with correct asymptotic behavior" *Phys. Rev. A* **38**, 3098 (1988).

[110] C. Lee, W. Yang, and R. G. Parr "Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density" *Phys. Rev. B* **37**, 785 (1988).

[111] L. Maschio, B. Civalleri, P. Ugliengo, and A. Gavezzotti "Intermolecular Interaction Energies in Molecular Crystals: Comparison and Agreement of Localized MøllerPlesset 2, Dispersion-Corrected Density Functional, and Classical Empirical Two-Body Calculations" *J. Phys. Chem. A* **115**, 11179 (2011).

[112] A. D. Becke "Density-functional thermochemistry. III. The role of exact exchange" *J. Chem. Phys.* **98**, 5648 (1993).

[113] S. H. Vosko, L. Wilk, and M. Nusair "Accurate spin-dependent electron liquid correlation energies for local spin density calculations: a critical analysis" *Can. J. Phys.* **58**, 1200 (1980).

[114] J. P. M. Trusler "Equation of State for Solid Phase I of Carbon Dioxide Valid for Temperatures up to 800 K and Pressures up to 12 GPa" *J. Phys. Chem. Ref. Data* **40**, 043105 (2011).

[115] A. Warshel and S. Lifson "Consistent Force Field Calculations. II. Crystal Structures, Sublimation

Energies, Molecular and Lattice Vibrations, Molecular Conformations, and Enthalpies of Alkanes" *J. Chem. Phys.* **53**, 582 (1970).

[116] H. Stoll and K. Doll "Approaching the bulk limit with finite cluster calculations using local increments: The case of LiH" *J. Chem. Phys.* **136**, 074106 (2012).

[117] D. Usvyat, B. Civalleri, L. Maschio, R. Dovesi, C. Pisani, and M. Schütz "Approaching the theoretical limit in periodic local MP2 calculations with atomic-orbital basis sets: The case of LiH" *J. Chem. Phys.* **134**, 214105 (2011).

[118] S. J. Nolan, M. J. Gillan, D. Alfè, N. L. Allan, and F. R. Manby "Calculation of properties of crystalline lithium hydride using correlated wave function theory" *Phys. Rev. B* **80**, 165109 (2009).

[119] J. J. Shepherd, A. Grueneis, G. H. Booth, G. Kresse, and A. Alavi "Convergence of many-body wave-function expansions using a plane-wave basis: From homogeneous electron gas to solid state systems" *Phys. Rev. B* **86**, 035111 (2012).

[120] A. Halkier, T. Helgaker, P. Jørgensen, W. Klopper, H. Koch, J. Olsen, and A. K. Wilson "Basis-set convergence in correlated calculations on Ne, N2, and H2O" *Chem. Phys. Lett.* **286**, 243 (1998).

[121] L. L. Shipman, A. W. Burgess, and H. A. Scheraga "Lattice energies and heats of sublimation at 0.deg.K for n-pentane, n-hexane, n-octane, and ammonia" *J. Phys. Chem.* **80**, 52 (1976).

[122] W. Keesom and J. Kohler "The lattice constant and expansion coefficient of solid carbon dioxide" *Physica* **1**, 655 (1934).

[123] A. Curzon "A comment on the lattice parameter of solid carbon dioxide at -190C" *Physica* **59**, 733 (1972).

[124] S. Grimme, J. Antony, S. Ehrlich, and H. Krieg "A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu" *J. Chem. Phys.* **132**, 154104 (2010).

[125] S. Grimme "Semiempirical GGA-type density functional constructed with a long-range dispersion correction" *J. Comput. Chem.* **27**, 1787 (2006).

[126] S. Grimme "Improved second-order Møller–Plesset perturbation theory by separate scaling of parallel- and antiparallel-spin pair correlation energies" *J. Chem. Phys.* **118**, 9095 (2003).

[127] Y. Jung, R. C. Lochan, A. D. Dutoi, and M. Head-Gordon "Scaled opposite-spin second order Møller–Plesset correlation energy: An economical electronic structure method" *J. Chem. Phys.* **121**, 9793 (2004).

[128] R. A. Distasio JR. and M. Head-Gordon "Optimized spin-component scaled second-order Møller-Plesset perturbation theory for intermolecular interaction energies" *Mol. Phys.* **105**, 1073 (2007).

[129] J. G. Hill and J. A. Platts "Spin-Component Scaling Methods for Weak and Stacking Interactions" *J. Chem. Theory Comput.* **3**, 80 (2007).

[130] P. Jurecka, J. Sponer, J. Cerny, and P. Hobza "Benchmark database of accurate (MP2 and CCSD(T) complete basis set limit) interaction energies of small model complexes, DNA base pairs, and amino acid pairs" *Phys. Chem. Chem. Phys.* **8**, 1985 (2006).

[131] M. Del Ben, J. Hutter, and J. VandeVondele "Electron Correlation in the Condensed Phase from a Resolution of Identity Approach Based on the Gaussian and Plane Waves Scheme" *J. Chem. Theory Comput.* **9**, 2654 (2013).

[132] Y. Jung, Y. Shao, and M. Head-Gordon "Fast evaluation of scaled opposite spin second-order Møller-Plesset correlation energies using auxiliary basis expansions and exploiting sparsity" *J. Comput. Chem.* **28**, 1953 (2007).

[133] H. Eshuis and F. Furche "Basis set convergence of molecular correlation energy differences within the random phase approximation" *J. Chem. Phys.* **136**, 084105 (2012).

[134] J. J. Shepherd, A. Grüneis, G. H. Booth, G. Kresse, and A. Alavi "Convergence of many-body wave-function expansions using a plane-wave basis: From homogeneous electron gas to solid state systems" *Phys. Rev. B* **86**, 035111 (2012).

[135] J. Toulouse, I. C. Gerber, G. Jansen, A. Savin, and J. G. Ángyán "Adiabatic-Connection Fluctuation-Dissipation Density-Functional Theory Based on Range Separation" *Phys. Rev. Lett.* **102**, 096404 (2009).

[136] I. C. Gerber and J. G. Angyan "London dispersion forces by range-separated hybrid density

functional with second order perturbational corrections: The case of rare gas complexes" *J. Chem. Phys.* **126**, 044103 (2007).

[137] K. Eichkorn, O. Treutler, H. Öhm, M. Häser, and R. Ahlrichs "Auxiliary basis sets to approximate Coulomb potentials" *Chem. Phys. Lett.* **240**, 283 (1995).

[138] X. Ren, P. Rinke, V. Blum, J. Wieferink, A. Tkatchenko, A. Sanfilippo, K. Reuter, and M. Scheffler "Resolution-of-identity approach to HartreeFock, hybrid density functionals, RPA, MP2 and GW with numeric atom-centered orbital basis functions" *New J. Phys.* **14**, 053020 (2012).

[139] D. E. Bernholdt and R. J. Harrison "Large-scale correlated electronic structure calculations: the RI-MP2 method on parallel computers" *Chem. Phys. Lett.* **250**, 477 (1996).

[140] L. Maschio and D. Usvyat "Fitting of local densities in periodic systems" *Phys. Rev. B* **78**, 073102 (2008).

[141] A. M. Burow, M. Sierka, and F. Mohamed "Resolution of identity approximation for the Coulomb term in molecular and periodic systems" *J. Chem. Phys.* **131**, 214101 (2009).

[142] M. Challacombe, C. White, and M. Head-Gordon "Periodic boundary conditions and the fast multipole method" *J. Chem. Phys.* **107**, 10131 (1997).

[143] K. N. Kudin and G. E. Scuseria "Revisiting infinite lattice sums with the periodic fast multipole method" *J. Chem. Phys.* **121**, 2886 (2004).

[144] J. L. Whitten "Coulombic potential energy integrals and approximations" *J. Chem. Phys.* **58**, 4496 (1973).

[145] B. I. Dunlap, J. W. D. Connolly, and J. R. Sabin "On some approximations in applications of X alpha theory" *J. Chem. Phys.* **71**, 3396 (1979).

[146] O. Vahtras, J. Almlöf, and M. Feyereisen "Integral approximations for LCAO-SCF calculations" *Chem. Phys. Lett.* **213**, 514 (1993).

[147] F. Weigend and M. Häser "RI-MP2: first derivatives and global consistency" *Theor. Chem. Acc.* **97**, 331 (1997).

[148] M. Head-Gordon, J. A. Pople, and M. J. Frisch "MP2 energy evaluation by direct methods" *Chem. Phys. Lett.* **153**, 503 (1988).

[149] A. Takatsuka, S. Ten-no, and W. Hackbusch "Minimax approximation for the decomposition of energy denominators in Laplace-transformed Møller–Plesset perturbation theories" *J. Chem. Phys.* **129**, 044112 (2008).

[150] D. Braess and W. Hackbusch "Approximation of 1/x by exponential sums in $[1, \infty)$" *IMA J. Numer. Anal.* **25**, 685 (2005).

[151] R. Bauernschmitt and R. Ahlrichs "Treatment of electronic excitations within the adiabatic approximation of time dependent density functional theory" *Chem. Phys. Lett.* **256**, 454 (1996).

[152] J. P. Boyd "Exponentially convergent Fourier/Chebyshev quadrature schemes on bounded and infinite intervals" *J. Sci. Comput.* **2**, 99 (1987).

[153] J. Ihm, A. Zunger, and M. L. Cohen "Momentum-space formalism for the total energy of solids" *J. Phys. C* **12**, 4409 (1979).

[154] T. Nakajima and K. Hirao "An approximate second-order Møller-Plesset perturbation approach for large molecular calculations" *Chem. Phys. Lett.* **427**, 225 (2006).

[155] D. Kats, D. Usvyat, S. Loibl, T. Merz, and M. Schütz "Comment on "Minimax approximation for the decomposition of energy denominators in Laplace-transformed Møller–Plesset perturbation theories" [J. Chem. Phys. **129**, 044112 (2008)]" *J. Chem. Phys.* **130**, 127101 (2009).

[156] F. Aryasetiawan, T. Miyake, and K. Terakura "Total Energy Method from Many-Body Formulation" *Phys. Rev. Lett.* **88**, 166401 (2002).

[157] H.-V. Nguyen and S. de Gironcoli "Efficient calculation of exact exchange and RPA correlation energies in the adiabatic-connection fluctuation-dissipation theory" *Phys. Rev. B* **79**, 205114 (2009).

[158] T. Miyake, F. Aryasetiawan, T. Kotani, M. van Schilfgaarde, M. Usuda, and K. Terakura "Total energy of solids: An exchange and random-phase approximation correlation study" *Phys. Rev. B* **66**, 245103 (2002).

[159] P. García-González, J. J. Fernández, A. Marini, and A. Rubio "Advanced Correlation Functionals:Application to Bulk Materials and Localized Systems" *J. Phys. Chem. A* **111**, 12458 (2007).

[160] J. Harl and G. Kresse "Accurate Bulk Properties from Approximate Many-Body Techniques" *Phys. Rev. Lett.* **103**, 056401 (2009).

[161] J. Harl, L. Schimka, and G. Kresse "Assessing the quality of the random phase approximation for lattice constants and atomization energies of solids" *Phys. Rev. B* **81**, 115126 (2010).

[162] B. Xiao, J. Sun, A. Ruzsinszky, J. Feng, and J. P. Perdew "Structural phase transitions in Si and SiO_2 crystals via the random phase approximation" *Phys. Rev. B* **86**, 094109 (2012).

[163] M. Rohlfing and T. Bredow "Binding Energy of Adsorbates on a Noble-Metal Surface: Exchange and Correlation Effects" *Phys. Rev. Lett.* **101**, 266106 (2008).

[164] X. Ren, P. Rinke, and M. Scheffler "Exploring the random phase approximation: Application to CO adsorbed on Cu(111)" *Phys. Rev. B* **80**, 045402 (2009).

[165] A. Marini, P. García-González, and A. Rubio "First-Principles Description of Correlation Effects in Layered Materials" *Phys. Rev. Lett.* **96**, 136404 (2006).

[166] F. Mittendorfer, A. Garhofer, J. Redinger, J. Klimeš, J. Harl, and G. Kresse "Graphene on Ni(111): Strong interaction and weak adsorption" *Phys. Rev. B* **84**, 201401 (2011).

[167] J. Harl and G. Kresse "Cohesive energy curves for noble gas solids calculated by adiabatic connection fluctuation-dissipation theory" *Phys. Rev. B* **77**, 045136 (2008).

[168] D. Lu, Y. Li, D. Rocca, and G. Galli "*Ab initio* Calculation of van der Waals Bonded Molecular Crystals" *Phys. Rev. Lett.* **102**, 206411 (2009).

[169] Y. Li, D. Lu, H.-V. Nguyen, and G. Galli "van der Waals Interactions in Molecular Assemblies from First-Principles Calculations" *J. Phys. Chem. A* **114**, 1944 (2010).

[170] J. P. Perdew, K. Burke, and M. Ernzerhof "Generalized Gradient Approximation Made Simple" *Phys. Rev. Lett.* **77**, 3865 (1996).

[171] T. Takatani, E. G. Hohenstein, M. Malagoli, M. S. Marshall, and C. D. Sherrill "Basis set consistent revision of the S22 test set of noncovalent interaction energies" *J. Chem. Phys.* **132**, 144104 (2010).

[172] J. F. Dobson and T. Gould "Calculation of dispersion energies" *J. Phys.: Condens. Matter* **24**, 073201 (2012).

[173] C. Adamo and V. Barone "Toward reliable density functional methods without adjustable parameters: The PBE0 model" *J. Chem. Phys.* **110**, 6158 (1999).

[174] M. Del Ben, O. Schütt, T. Wentz, P. Messmer, J. Hutter, and J. VandeVondele "Enabling simulation at the fifth rung of DFT: Large scale RPA calculations with excellent time to solution" *Comput. Phys. Commun.* **187**, 120 (2015).

[175] M. J. McGrath, J. I. Siepmann, I.-F. W. Kuo, C. J. Mundy, J. VandeVondele, J. Hutter, F. Mohamed, and M. Krack "Isobaric–Isothermal Monte Carlo Simulations from First Principles: Application to Liquid Water at Ambient Conditions" *ChemPhysChem* **6**, 1894 (2005).

[176] J. Schmidt, J. VandeVondele, I.-F. W. Kuo, D. Sebastiani, J. I. Siepmann, J. Hutter, and C. J. Mundy "Isobaric–Isothermal Molecular Dynamics Simulations Utilizing Density Functional Theory: An Assessment of the Structure and Density of Water at Near–Ambient Conditions" *J. Phys. Chem. B* **113**, 11959 (2009).

[177] M. Del Ben, M. Schönherr, J. Hutter, and J. VandeVondele "Bulk Liquid Water at Ambient Temperature and Pressure from MP2 Theory" *J. Phys. Chem. Lett.* **4**, 3753 (2013).

[178] L. Schimka, J. Harl, A. Stroppa, A. Grüneis, M. Marsman, F. Mittendorfer, and G. Kresse "Accurate surface and adsorption energies from many-body perturbation theory" *Nature Mat.* **9**, 741 (2010).

[179] M. Kaltak, J. Klimeš, and G. Kresse "Low Scaling Algorithms for the Random Phase Approximation: Imaginary Time and Laplace Transformations" *J. Chem. Theory Comput.* **10**, 2498 (2014).

[180] O. Schütt, P. Messmer, J. Hutter, and J. VandeVondele *GPU Accelerated Sparse Matrix Matrix Multiplication for Linear Scaling Density Functional Theory* (2014).

[181] T. Olsen and K. S. Thygesen "Random phase approximation applied to solids, molecules, and graphene-metal interfaces: From van der Waals to covalent bonding" *Phys. Rev. B* **87**, 075111 (2013).

[182] L. Schimka, R. Gaudoin, J. c. v. Klimeš, M. Marsman, and G. Kresse "Lattice constants and cohesive energies of alkali, alkaline-earth, and transition metals: Random phase approximation and density functional theory results" *Phys. Rev. B* **87**, 214102 (2013).

[183] M. Macher, J. Klimes, C. Franchini, and G. Kresse "The random phase approximation applied to ice" *J. Chem. Phys.* **140**, 084502 (2014).

[184] D. E. Bernholdt "Scalability of correlated electronic structure calculations on parallel computers: A case study of the RI-MP2 method" *Parallel Comput.* **26**, 945 (2000).

[185] M. Katouda and T. Nakajima "MPI/OpenMP Hybrid Parallel Algorithm of Resolution of Identity Second-Order Møller-Plesset Perturbation Calculation for Massively Parallel Multicore Supercomputers" *J. Chem. Theory Comput.* **9**, 5373 (2013).

[186] A. M. Burow, J. E. Bates, F. Furche, and H. Eshuis "Analytical First-Order Molecular Properties and Forces within the Adiabatic Connection Random Phase Approximation" *J. Chem. Theory Comput.* **10**, 180 (2014).

[187] D. Braess *Nonlinear Approximation Theory* Springer Series in Computational Mathematics. Springer London, Limited (2011).

[188] E. W. Lang and H. D. Lüdemann "Anomalies of Liquid Water" *Angew. Chem. Int. Ed. Engl.* **21**, 315 (1982).

[189] C. A. Angell "Insights into Phases of Liquid Water from Study of Its Unusual Glass-Forming Properties" *Science* **319**, 582 (2008).

[190] C. A. Tulk, C. J. Benmore, J. Urquidi, D. D. Klug, J. Neuefeind, B. Tomberli, and P. A. Egelstaff "Structural Studies of Several Distinct Metastable Forms of Amorphous Ice" *Science* **297**, 1320 (2002).

[191] P. H. Poole, F. Sciortino, U. Essmann, and H. E. Stanley "Phase behaviour of metastable water" *Nature* **360**, 324 (1992).

[192] M. Watkins, D. Pan, E. G. Wang, A. Michaelides, J. VandeVondele, and B. Slater "Large variation of vacancy formation energies in the surface of crystalline ice" *Nat. Mater.* **10**, 794 (2011).

[193] J. R. Errington and P. G. Debenedetti "Relationship between structural order and the anomalies of liquid water" *Nature* **409**, 318 (2001).

[194] O. Mishima1 and H. E. Stanley "The relationship between liquid, supercooled and glassy water" *Nature* **396**, 329 (1998).

[195] P. Wernet, D. Nordlund, U. Bergmann, M. Cavalleri, M. Odelius, H. Ogasawara, L. A. Näslund, T. K. Hirsch, L. Ojamäe, P. Glatzel, L. G. M. Pettersson, and A. Nilsson "The Structure of the First Coordination Shell in Liquid Water" *Science* **304**, 995 (2004).

[196] C. Huang, K. T. Wikfeldt, T. Tokushima, D. Nordlund, Y. Harada, U. Bergmann, M. Niebuhr, T. M. Weiss, Y. Horikawa, M. Leetmaa, M. P. Ljungberg, O. Takahashi, A. Lenz, L. Ojamäe, A. P. Lyubartsev, S. Shin, L. G. M. Pettersson, and A. Nilsson "The inhomogeneous structure of water at ambient conditions" *Proc. Natl. Acad. Sci.* **106**, 15214 (2009).

[197] J. D. Smith, C. D. Cappa, K. R. Wilson, B. M. Messer, R. C. Cohen, and R. J. Saykally "Energetics of Hydrogen Bond Network Rearrangements in Liquid Water" *Science* **306**, 851 (2004).

[198] I. Ohmine and H. Tanaka "Fluctuation, relaxations, and hydration in liquid water. Hydrogen-bond rearrangement dynamics" *Chem. Rev.* **93**, 2545 (1993).

[199] K. Stokely, M. G. Mazza, H. E. Stanley, and G. Franzese "Effect of hydrogen bond cooperativity on the behavior of water" *Proc. Natl. Acad. Sci.* **107**, 1301 (2010).

[200] Chaplin M. http://www1.lsbu.ac.uk/water/anmlies.html (2014).

[201] A. Rahman and F. H. Stillinger "Molecular Dynamics Study of Liquid Water" *J. Chem. Phys.* **55**, 3336 (1971).

[202] K. Laasonen, M. Sprik, M. Parrinello, and R. Car ""Ab Initio" Liquid Water" *J. Chem. Phys.* **99**, 9080 (1993).

[203] J. VandeVondele, F. Mohamed, M. Krack, J. Hutter, M. Sprik, and M. Parrinello "The influence of temperature and density functional models in ab initio molecular dynamics simulation of liquid water" *J. Chem. Phys.* **122**, 014515 (2005).

[204] M. Schönherr, B. Slater, J. Hutter, and J. VandeVondele "Dielectric Properties of Water Ice, the Ice Ih/XI Phase Transition, and an Assessment of Density Functional Theory" *J. Phys. Chem. B* **118**, 590 (2014).

[205] C. Adriaanse, J. Cheng, V. Chau, M. Sulpizi, J. VandeVondele, and M. Sprik "Aqueous Redox

Chemistry and the Electronic Band Structure of Liquid Water" *J. Phys. Chem. Lett.* **3**, 3411 (2012).

[206] M. Del Ben, J. Hutter, and J. VandeVondele "In preparation" (2014).

[207] J. Hutter, M. Iannuzzi, F. Schiffmann, and J. VandeVondele "CP2K: atomistic simulations of condensed matter systems" *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **4**, 15 (2014).

[208] V. Weber, V. Bekas, T. Laino, and A. Curioni "Shedding Light on Lithium/Air Batteries Using Millions of Threads on the BG/Q Supercomputer" *http://cpmd.org/the-code/performance-scale-out-data/PID3080165.pdf* (2014).

[209] T. Auckenthaler, V. Blum, H. J. Bungartz, T. Huckle, R. Johanni, L. Krämer, B. Lang, H. Lederer, and P. R. Willems "Parallel Solution of Partial Symmetric Eigenvalue Problems from Electronic Structure Calculations" *Parallel Comput.* **37**, 783 (2011).

[210] A. Haidar, R. Solcà, M. Gates, S. Tomov, T. Schulthess, and J. Dongarra "Leading edge hybrid multi-GPU algorithms for generalized eigenproblems in electronic structure calculations" *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* **7905 LNCS**, 67 (2013).

[211] M. Del Ben, J. Hutter, and J. VandeVondele "Force and Stress in Second Order Møller-Plesset Perturbation Theory for Condensed Phase Systems within the Resolution-of-Identity Gaussian and Plane Waves Approach" *J. Chem. Phys. Accepted* (2015).

[212] S. Y. Willow, K. S. Kim, and S. Hirata "Stochastic evaluation of second-order many-body perturbation energies" *J. Chem. Phys.* **137**, 204122 (2012).

[213] S. Y. Willow, M. R. Hermes, K. S. Kim, and S. Hirata "Convergence Acceleration of Parallel Monte Carlo Second-Order Many-Body Perturbation Calculations Using Redundant Walkers" *J. Chem. Theory Comput.* **9**, 4396 (2013).

[214] D. Neuhauser, E. Rabani, and R. Baer "Expeditious Stochastic Approach for MP2 Energies in Large Electronic Systems" *J. Chem. Theory Comput.* **9**, 24 (2013).

[215] Q. Ge, Y. Gao, R. Baer, E. Rabani, and D. Neuhauser "A Guided Stochastic Energy-Domain Formulation of the Second Order Møller-Plesset Perturbation Theory" *J. Phys. Chem. Lett.* **5**, 185 (2014).

[216] C. Hättig, W. Klopper, A. Köhn, and D. P. Tew "Explicitly Correlated Electrons in Molecules" *Chem. Rev.* **112**, 4 (2012).

[217] A. Grüneis, J. J. Shepherd, A. Alavi, D. P. Tew, and G. H. Booth "Explicitly correlated plane waves: Accelerating convergence in periodic wavefunction expansions" *J. Chem. Phys.* **139**, 084112 (2013).

[218] F. Weigend, A. Köhn, and C. Hättig "Efficient use of the correlation consistent basis sets in resolution of the identity MP2 calculations" *J. Chem. Phys.* **116**, 3175 (2002).

[219] J. A. Pople, R. Krishnan, H. B. Schlegel, and J. S. Binkley "Derivative studies in Hartree-Fock and Møller-Plesset theories" *Int. J. Quantum Chem. Symp.* **13**, 225 (1979).

[220] E. Kraka, J. Gauss, and D. Cremer "Determination and use of response densities" *J. Mol. Struct.: THEOCHEM* **234**, 95 (1991).

[221] T. Helgaker, P. Jørgensen, and N. Handy "A numerically stable procedure for calculating Møller-Plesset energy derivatives, derived using the theory of Lagrangians" *Theor. Chim. Acta* **76**, 227 (1989).

[222] C. M. Aikens, S. P. Webb, R. L. Bell, G. D. Fletcher, M. W. Schmidt, and M. S. Gordon "A derivation of the frozen-orbital unrestricted open-shell and restricted closed-shell second-order perturbation theory analytic gradient expressions" *Theor. Chem. Acc.* **110**, 233 (2003).

[223] M. J. Frisch, M. Head-Gordon, and J. A. Pople "A direct MP2 gradient method" *Chem. Phys. Lett.* **166**, 275 (1990).

[224] M. Head-Gordon "An improved semidirect MP2 gradient method" *Mol. Phys.* **96**, 673 (1999).

[225] R. A. Distasio, R. P. Steele, Y. M. Rhee, Y. Shao, and M. Head-Gordon "An improved algorithm for analytical gradient evaluation in resolution-of-the-identity second-order Møller-Plesset perturbation theory: Application to alanine tetrapeptide conformational analysis" *J. Comput. Chem.* **28**, 839 (2007).

[226] I. M. Nielsen "A new direct MP2 gradient algorithm with implementation on a massively parallel

computer" *Chem. Phys. Lett.* **255**, 210 (1996).

[227] G. D. Fletcher, A. P. Rendell, and P. Sherwood "A parallel second-order Møller-Plesset gradient" *Mol. Phys.* **91**, 431 (1997).

[228] K. Ishimura, P. Pulay, and S. Nagase "New parallel algorithm for MP2 energy gradient calculations" *J. Comput. Chem.* **28**, 2034 (2007).

[229] C. Hättig, A. Hellweg, and A. Köhn "Distributed memory parallel implementation of energies and gradients for second-order Møller-Plesset perturbation theory with the resolution-of-the-identity approximation" *Phys. Chem. Chem. Phys.* **8**, 1159 (2006).

[230] J. Gerratt and I. M. Mills "Force Constants and DipoleMoment Derivatives of Molecules from Perturbed Hartree-Fock Calculations. I" *J. Chem. Phys.* **49**, 1719 (1968).

[231] N. C. Handy and H. F. Schaefer "On the evaluation of analytic energy derivatives for correlated wave functions" *J. Chem. Phys.* **81**, 5031 (1984).

[232] T. J. Lee, S. C. Racine, J. E. Rice, and A. P. Rendell "On the orbital contribution to analytical derivatives of perturbation theory energies" *Mol. Phys.* **85**, 561 (1995).

[233] Y. Osamura, Y. Yamaguchi, P. Saxe, D. Fox, M. Vincent, and H. Schaefer "Analytic second derivative techniques for self-consistent-field wave functions. A new approach to the solution of the coupled perturbed Hartree-Fock equations" *J. Mol. Struct.: THEOCHEM* **103**, 183 (1983).

[234] V. Weber and C. Daul "Improved coupled perturbed Hartree-Fock and Kohn-Sham convergence acceleration" *Chem. Phys. Lett.* **370**, 99 (2003).

[235] T. Abe, Y. Sekine, and F. Sato "Solving a coupled perturbed equation by the residual cutting method" *Chem. Phys. Lett.* **557**, 176 (2013).

[236] H. Weiss, R. Ahlrichs, and M. Häser "A direct algorithm for selfconsistentfield linear response theory and application to C60: Excitation energies, oscillator strengths, and frequencydependent polarizabilities" *J. Chem. Phys.* **99**, 1262 (1993).

[237] O. H. Nielsen and R. M. Martin "First-Principles Calculation of Stress" *Phys. Rev. Lett.* **50**, 697 (1983).

[238] O. H. Nielsen and R. M. Martin "Quantum-mechanical theory of stress and force" *Phys. Rev. B* **32**, 3780 (1985).

[239] D. Marx and J. Hutter "Ab initio molecular dynamics: Theory and implementation" in "Modern Methods and Algorithms of Quantum Chemistry, John Von Neumann Institute for Computing, Forschungszentrum Julich, 2000", pp. 329–477 (2000).

[240] A. Dal Corso and R. Resta "Density-functional theory of macroscopic stress: Gradient-corrected calculations for crystalline Se" *Phys. Rev. B* **50**, 4327 (1994).

[241] L. C. Balbás, J. L. Martins, and J. M. Soler "Evaluation of exchange-correlation energy, potential, and stress" *Phys. Rev. B* **64**, 165110 (2001).

[242] K. Doll, R. Dovesi, and R. Orlando "Analytical Hartree-Fock gradients with respect to the cell parameter for systems periodic in three dimensions" *Theor. Chem. Acc.* **112**, 394 (2004).

[243] A. W. Hewat and C. Riekel "The crystal structure of deuteroammonia between 2 and 180 K by neutron powder profile refinement" *Acta Crystallogr. Sect. A* **35**, 569 (1979).

[244] R. Boese, N. Niederprüm, D. Bläser, A. Maulitz, M. Y. Antipin, and P. R. Mallinson "Single-Crystal Structure and Electron Density Distribution of Ammonia at 160 K on the Basis of X-ray Diffraction Data" *J. Phys. Chem. B* **101**, 5794 (1997).

[245] A. Simon and K. Peters "Single-crystal refinement of the structure of carbon dioxide" *Acta Crystallogr. Sect. B* **36**, 2750 (1980).

[246] C. Müller and D. Usvyat "Incrementally Corrected Periodic Local MP2 Calculations: I. The Cohesive Energy of Molecular Crystals" *J. Chem. Theory Comput.* **9**, 5590 (2013).

[247] P. J. Bygrave, N. L. Allan, and F. R. Manby "The embedded many-body expansion for energetics of molecular crystals" *J. Chem. Phys.* **137**, 164102 (2012).

[248] O. Sode, M. Keceli, K. Yagi, and S. Hirata "Fermi resonance in solid CO2 under pressure" *J. Chem. Phys.* **138**, 074501 (2013).

[249] S. Wen and G. J. O. Beran "Accurate Molecular Crystal Lattice Energies from a Fragment QM/MM Approach with On-the-Fly Ab Initio Force Field Parametrization" *J. Chem. Theory Comput.* **7**,

3733 (2011).

[250] S. Tosoni, C. Tuma, J. Sauer, B. Civalleri, and P. Ugliengo "A comparison between plane wave and Gaussian-type orbital basis sets for hydrogen bonded systems: Formic acid as a test case" *J. Chem. Phys.* **127**, 154102 (2007).

[251] S. Hirata "Fast electron-correlation methods for molecular crystals: An application to the alpha, beta1, and beta2 modifications of solid formic acid" *J. Chem. Phys.* **129**, 204104 (2008).

[252] K. E. Riley and P. Hobza "Assessment of the MP2 Method, along with Several Basis Sets, for the Computation of Interaction Energies of Biologically Relevant Hydrogen Bonded and Dispersion Bound Complexes" *J. Phys. Chem. A* **111**, 8257 (2007).

[253] F. Neese, T. Schwabe, and S. Grimme "Analytic derivatives for perturbatively corrected double hybrid density functionals: Theory, implementation, and applications" *J. Chem. Phys.* **126**, 124115 (2007).

[254] R. C. Lochan, Y. Shao, and M. Head-Gordon "Quartic-Scaling Analytical Energy Gradient of Scaled Opposite-Spin Second-Order Møller-Plesset Perturbation Theory" *J. Chem. Theory Comput.* **3**, 988 (2007).

[255] H. Ji, Y. Shao, W. A. Goddard, and Y. Jung "Analytic Derivatives of Quartic-Scaling Doubly Hybrid XYGJ-OS Functional: Theory, Implementation, and Benchmark Comparison with M06-2X and MP2 Geometries for Nonbonded Complexes" *J. Chem. Theory Comput.* **9**, 1971 (2013).

[256] M. Del Ben, M. Schönherr, J. Hutter, and J. VandeVondele "Correction to "Bulk Liquid Water at Ambient Temperature and Pressure from MP2 Theory"" *J. Phys. Chem. Lett.* **5**, 3066 (2014).

[257] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein "Comparison of simple potential functions for simulating liquid water" *J. Chem. Phys.* **79**, 926 (1983).

[258] B. Guillot "A reappraisal of what we have learnt during three decades of computer simulations on water" *J. Mol. Liq.* **101**, 219 (2002).

[259] W. L. Jorgensen and J. Tirado-Rives "Potential energy functions for atomic-level simulations of water and organic and biomolecular systems" *Proc. Natl. Acad. Sci. U.S.A.* **102**, 6665 (2005).

[260] R. Bukowski, K. Szalewicz, G. C. Groenenboom, and A. van der Avoird "Predictions of the Properties of Water from First Principles" *Science* **315**, 1249 (2007).

[261] A. G. Donchev, N. G. Galkin, A. A. Illarionov, O. V. Khoruzhii, M. A. Olevanov, V. D. Ozrin, M. V. Subbotin, and V. I. Tarasov "Water properties from first principles: Simulations by a general-purpose quantum mechanical polarizable force field" *Proc. Natl. Acad. Sci. U.S.A.* **103**, 8613 (2006).

[262] M. W. Mahoney and W. L. Jorgensen "A five-site model for liquid water and the reproduction of the density anomaly by rigid, nonpolarizable potential functions" *J. Chem. Phys.* **112**, 8910 (2000).

[263] S. W. Rick, S. J. Stuart, and B. J. Berne "Dynamical fluctuating charge force fields: Application to liquid water" *J. Chem. Phys.* **101**, 6141 (1994).

[264] P. Paricaud, M. Predota, A. A. Chialvo, and P. T. Cummings "From dimer to condensed phases at extreme conditions: Accurate predictions of the properties of water by a Gaussian charge polarizable model" *J. Chem. Phys.* **122**, 244511 (2005).

[265] B. Chen, J. Xing, and J. I. Siepmann "Development of Polarizable Water Force Fields for Phase Equilibrium Calculations" *J. Phys. Chem. B* **104**, 2391 (2000).

[266] T. M. Truskett and K. A. Dill "A Simple Statistical Mechanical Model of Water" *J. Phys. Chem. B* **106**, 11829 (2002).

[267] H. Stanley, S. V. Buldyrev, and N. Giovambattista "Static heterogeneities in liquid water" *Physica A* **342**, 40 (2004).

[268] A. Nilsson and L. Pettersson "Perspective on the structure of liquid water" *Chemical Physics* **389**, 1 (2011).

[269] I.-F. W. Kuo, C. J. Mundy, M. J. McGrath, J. I. Siepmann, J. VandeVondele, M. Sprik, J. Hutter, B. Chen, M. L. Klein, F. Mohamed, M. Krack, and M. Parrinello "Liquid Water from First Principles: Investigation of Different Sampling Approaches" *J. Phys. Chem. B* **108**, 12990 (2004).

[270] P. L. Silvestrelli and M. Parrinello "Structural, electronic, and bonding properties of liquid water from first principles" *J. Chem. Phys.* **111**, 3572 (1999).

[271] S. Izvekov and G. A. Voth "Car–Parrinello molecular dynamics simulation of liquid water: New results" *J. Chem. Phys.* **116**, 10372 (2002).

[272] J. C. Grossman, E. Schwegler, E. W. Draeger, F. Gygi, and G. Galli "Towards an assessment of the accuracy of density functional theory for first principles simulations of water" *J. Chem. Phys.* **120**, 300 (2004).

[273] T. D. Kühne, M. Krack, and M. Parrinello "Static and Dynamical Properties of Liquid Water from First Principles by a Novel Car–Parrinello–like Approach" *J. Chem. Theory Comput.* **5**, 235 (2009).

[274] I.-C. Lin, A. P. Seitsonen, I. Tavernelli, and U. Rothlisberger "Structure and Dynamics of Liquid Water from ab Initio Molecular Dynamics–Comparison of BLYP, PBE, and revPBE Density Functionals with and without van der Waals Corrections" *J. Chem. Theory Comput.* **8**, 3902 (2012).

[275] E. Schwegler, J. C. Grossman, F. Gygi, and G. Galli "Towards an assessment of the accuracy of density functional theory for first principles simulations of water. II" *J. Chem. Phys.* **121**, 5400 (2004).

[276] R. Jonchiere, A. P. Seitsonen, G. Ferlat, A. M. Saitta, and R. Vuilleumier "Van der Waals effects in ab initio water at ambient and supercritical conditions" *J. Chem. Phys.* **135**, 154503 (2011).

[277] A. Møgelhøj, A. K. Kelkkanen, K. T. Wikfeldt, J. Schiøtz, J. J. Mortensen, L. G. M. Pettersson, B. I. Lundqvist, K. W. Jacobsen, A. Nilsson, and J. K. Nørskov "Ab Initio van der Waals Interactions in Simulations of Water Alter Structure from Mainly Tetrahedral to High-Density-Like" *J. Phys. Chem. B* **115**, 14149 (2011).

[278] Z. Ma, Y. Zhang, and M. E. Tuckerman "Ab initio molecular dynamics study of water at constant pressure using converged basis sets and empirical dispersion corrections" *J. Chem. Phys.* **137**, 044506 (2012).

[279] S. Nosé and M. Klein "Constant pressure molecular dynamics for molecular systems" *Mol. Phys.* **50**, 1055 (1983).

[280] R. Iftimie, D. Salahub, D. Wei, and J. Schofield "Using a classical potential as an efficient importance function for sampling from an ab initio potential" *J. Chem. Phys.* **113**, 4852 (2000).

[281] B. Hetenyi, K. Bernacki, and B. J. Berne "Multiple "time step" Monte Carlo" *J. Chem. Phys.* **117**, 8203 (2002).

[282] L. D. Gelb "Monte Carlo simulations using sampling from an approximate potential" *J. Chem. Phys.* **118**, 7747 (2003).

[283] S. Duane, A. Kennedy, B. J. Pendleton, and D. Roweth "Hybrid Monte Carlo" *Phys. Lett. B* **195**, 216 (1987).

[284] B. Mehlig, D. W. Heermann, and B. M. Forrest "Hybrid Monte Carlo method for condensed-matter systems" *Phys. Rev. B* **45**, 679 (1992).

[285] S. Grimme "Accurate description of van der Waals complexes by density functional theory including empirical corrections" *J. Comput. Chem.* **25**, 1463 (2004).

[286] S. S. Xantheas, C. J. Burnham, and R. J. Harrison "Development of transferable interaction models for water. II. Accurate energetics of the first few water clusters from first principles" *J. Chem. Phys.* **116**, 1493 (2002).

[287] S. S. Xantheas and E. Apra "The binding energies of the D_{2d} and S_4 water octamer isomers: High–level electronic structure and empirical potential results" *J. Chem. Phys.* **120**, 823 (2004).

[288] I. Shin, M. Park, S. K. Min, E. C. Lee, S. B. Suh, and K. S. Kim "Structure and spectral features of $H^+(H_2O)_7$: Eigen versus Zundel forms" *J. Chem. Phys.* **125**, 234305 (2006).

[289] E. E. Dahlke, R. M. Olson, H. R. Leverentz, and D. G. Truhlar "Assessment of the Accuracy of Density Functionals for Prediction of Relative Energies and Geometries of Low-Lying Isomers of Water Hexamers" *J. Phys. Chem. A* **112**, 3976 (2008).

[290] L. Goerigk and S. Grimme "A thorough benchmark of density functional methods for general main group thermochemistry, kinetics, and noncovalent interactions" *Phys. Chem. Chem. Phys.* **13**, 6670 (2011).

[291] A. Erba, L. Maschio, S. Salustro, and S. Casassa "A post-Hartree–Fock study of pressure-induced phase transitions in solid nitrogen: The case of the alpha, gamma, and epsilon low-pressure phases" *J. Chem. Phys.* **134**, 074502 (2011).

[292] F. Goltl, A. Gruneis, T. Bucko, and J. Hafner "Van der Waals interactions between hydrocarbon molecules and zeolites: Periodic calculations at different levels of theory, from density functional theory to the random phase approximation and Møller-Plesset perturbation theory" *J. Chem. Phys.* **137**, 114111 (2012).

[293] M. Guidon, J. Hutter, and J. VandeVondele "Auxiliary Density Matrix Methods for Hartree–Fock Exchange Calculations" *J. Chem. Theory Comput.* **6**, 2348 (2010).

[294] M. J. McGrath, J. I. Siepmann, I.-F. W. Kuo, and C. J. Mundy "Vapor-liquid equilibria of water from first principles: comparison of density functionals and basis sets" *Mol. Phys.* **104**, 3619 (2006).

[295] S. Izvekov, M. Parrinello, C. J. Burnham, and G. A. Voth "Effective force fields for condensed phase systems from ab initio molecular dynamics simulation: A new method for force-matching" *J. Chem. Phys.* **120**, 10896 (2004).

[296] E. E. Dahlke and D. G. Truhlar "Improved Density Functionals for Water" *J. Phys. Chem. B* **109**, 15677 (2005).

[297] L. B. Skinner, C. Huang, D. Schlesinger, L. G. M. Pettersson, A. Nilsson, and C. J. Benmore "Benchmark oxygen-oxygen pair-distribution function of ambient water from x-ray diffraction measurements with a wide Q-range" *J. Chem. Phys.* **138**, 074506 (2013).

[298] R. A. Kuharski and P. J. Rossky "A quantum mechanical study of structure in liquid H_2O and D_2O" *J. Chem. Phys.* **82**, 5164 (1985).

[299] J. Lobaugh and G. A. Voth "A quantum model for water: Equilibrium and dynamical properties" *J. Chem. Phys.* **106**, 2400 (1997).

[300] M. W. Mahoney and W. L. Jorgensen "Quantum, intramolecular flexibility, and polarizability effects on the reproduction of the density anomaly of liquid water by simple potential functions" *J. Chem. Phys.* **115**, 10758 (2001).

[301] J. A. Morrone and R. Car "Nuclear Quantum Effects in Water" *Phys. Rev. Lett.* **101**, 017801 (2008).

[302] F. Paesani, S. Iuchi, and G. A. Voth "Quantum effects in liquid water from an ab initio-based polarizable force field" *J. Chem. Phys.* **127**, 074506 (2007).

[303] J. Wang, G. Román-Pérez, J. M. Soler, E. Artacho, and M.-V. Fernández-Serra "Density, structure, and dynamics of water: The effect of van der Waals interactions" *J. Chem. Phys.* **134**, 024516 (2011).

[304] C. Zhang, D. Donadio, F. Gygi, and G. Galli "First Principles Simulations of the Infrared Spectrum of Liquid Water Using Hybrid Density Functionals" *J. Chem. Theory Comput.* **7**, 1443 (2011).

[305] B. Santra, A. Michaelides, and M. Scheffler "On the accuracy of density-functional theory exchange-correlation functionals for H bonds in small water clusters: Benchmarks approaching the complete basis set limit" *J. Chem. Phys.* **127**, 184104 (2007).

[306] B. Santra, A. Michaelides, and M. Scheffler "Coupled cluster benchmarks of water monomers and dimers extracted from density-functional theory liquid water: The importance of monomer deformations" *J. Chem. Phys.* **131**, 124509 (2009).

[307] H.-S. Lee and M. E. Tuckerman "Structure of liquid water at ambient temperature from ab initio molecular dynamics performed in the complete basis set limit" *J. Chem. Phys.* **125**, 154507 (2006).

[308] M. Del Ben, J. VandeVondele, and B. Slater "Periodic MP2, RPA, and Boundary Condition Assessment of Hydrogen Ordering in Ice XV" *J. Phys. Chem. Lett.* **5**, 4122 (2014).

[309] K. D. Nanda and G. J. O. Beran "What Governs the Proton Ordering in Ice XV?" *J. Phys. Chem. Lett.* **4**, 3165 (2013).

[310] C. G. Salzmann, P. G. Radaelli, B. Slater, and J. L. Finney "The polymorphism of ice: five unresolved questions" *Phys. Chem. Chem. Phys.* **13**, 18468 (2011).

[311] T. L. Malkin, B. J. Murray, V. Andrey, J. Anwar, C. G. Salzmann, T. L. Malkin, B. J. Murray, A. V. Brukhno, J. Anwar, and C. G. Salzmann "Structure of ice crystallized from supercooled water" *Proc. Natl. Acad. Sci. U.S.A.* **109**, 1041 (2012).

[312] J. Russo, F. Romano, and H. Tanaka "New metastable form of ice and its role in the homogeneous crystallization of water." *Nat. Mater.* **13**, 733 (2014).

[313] B. Slater and D. Quigley "Crystal nucleation: Zeroing in on ice." *Nat. Mater.* **13**, 670 (2014).

[314] J. A. Sellberg, C. Huang, T. a. McQueen, N. D. Loh, H. Laksmono, D. Schlesinger, R. G. Sierra, D. Nordlund, C. Y. Hampton, D. Starodub, D. P. DePonte, M. Beye, C. Chen, a. V. Martin, a. Barty, K. T. Wikfeldt, T. M. Weiss, C. Caronna, J. Feldkamp, L. B. Skinner, M. M. Seibert, M. Messerschmidt, G. J. Williams, S. Boutet, L. G. M. Pettersson, M. J. Bogan, and a. Nilsson "Ultrafast X-ray probing of water structure below the homogeneous ice nucleation temperature" *Nature* **510**, 381 (2014).

[315] C. Salzmann, P. Radaelli, E. Mayer, and J. Finney "Ice XV: A New Thermodynamically Stable Phase of Ice" *Phys. Rev. Lett.* **103**, 105701 (2009).

[316] C. Knight and S. J. Singer "Prediction of a phase transition to a hydrogen bond ordered form of ice VI." *J. Phys. Chem. B* **109**, 21040 (2005).

[317] J.-L. Kuo "The low-temperature proton-ordered phases of ice predicted by ab initio methods" *Phys. Chem. Chem. Phys.* **5**, 3733 (2005).

[318] J.-L. Kuo and W. F. Kuhs "A First Principles Study on the Structure of Ice-VI: Static Distortion, Molecular Geometry, and Proton Ordering" *J. Phys. Chem. B* **110**, 3697 (2006).

[319] S. Singer, J.-L. Kuo, T. Hirsch, C. Knight, L. Ojamäe, and M. Klein "Hydrogen-Bond Topology and the Ice VII/VIII and Ice Ih/XI Proton-Ordering Phase Transitions" *Phys. Rev. Lett.* **94**, 135701 (2005).

[320] J.-L. Kuo, J. V. Coe, S. J. Singer, Y. B. Band, and L. Ojamäe "On the use of graph invariants for efficiently generating hydrogen bond topologies and predicting physical properties of water clusters and ice" *J. Chem. Phys.* **114**, 2527 (2001).

[321] Z. Raza, D. Alfe, C. G. Salzmann, J. Klimeš, A. Michaelides, and B. Slater "Proton ordering in cubic ice and hexagonal ice; a potential new ice phase-XIc" *Phys. Chem. Chem. Phys.* **13**, 19788 (2011).

[322] X. Fan, D. Bing, J. Zhang, Z. Shen, and J.-L. Kuo "Predicting the hydrogen bond ordered structures of ice Ih, II, III, VI and ice VII: DFT methods with localized based set" *Comput. Mater. Sci.* **49**, S170 (2010).

[323] S. Singer and C. Knight "Hydrogen bond ordering in ice V and the transition to ice XIII" *J. Chem. Phys.* **129**, 164513 (2008).

[324] J.-L. Kuo and M. L. Klein "Structure of Ice-VII and Ice-VIII: A Quantum Mechanical Study" *J. Phys. Chem. B* **108**, 19634 (2004).

[325] G. A. Tribello and B. Slater "Proton ordering energetics in ice phases" *Chem. Phys. Lett.* **425**, 246 (2006).

[326] S. J. Singer and C. Knight "Hydrogen-Bond Topology and Proton Ordering in Ice and Water Clusters" *Adv. Chem. Phys.* **147**, 1 (2011).

[327] B. Santra, J. Klimeš, D. Alfe, A. Tkatchenko, B. Slater, A. Michaelides, R. Car, and M. Scheffler "Hydrogen Bonds and van der Waals Forces in Ice at Ambient and High Pressures" *Phys. Rev. Lett.* **107** (2011).

[328] B. Santra, J. Klimeš, A. Tkatchenko, D. Alfè, B. Slater, A. Michaelides, R. Car, and M. Scheffler "On the accuracy of van der Waals inclusive density-functional theory exchange-correlation functionals for ice at ambient and high pressures." *J. Chem. Phys.* **139**, 154702 (2013).

[329] J. Klimeš and A. Michaelides "Perspective: Advances and challenges in treating van der Waals dispersion forces in density functional theory" *J. Chem. Phys.* **137**, 120901 (2012).

[330] R. a. DiStasio, V. V. Gobre, and A. Tkatchenko "Many-body van der Waals interactions in molecules and condensed matter." *J. Phys. Condens. Matter* **26**, 213202 (2014).

[331] A. P. Bartók, M. J. Gillan, F. R. Manby, and G. Csányi "Machine-learning approach for one- and two-body corrections to density functional theory: Applications to molecular and condensed water" *Phys. Rev. B* **88**, 054104 (2013).

[332] M. J. Gillan, D. Alfè, P. J. Bygrave, C. R. Taylor, and F. R. Manby "Energy benchmarks for water clusters and ice structures from an embedded many-body expansion." *J. Chem. Phys.* **139**, 114101 (2013).

[333] A. Hermann and P. Schwerdtfeger "Ground-State Properties of Crystalline Ice from Periodic Hartree-Fock Calculations and a Coupled-Cluster-Based Many-Body Decomposition of the Correlation

Energy" *Phys. Rev. Lett.* **101**, 183005 (2008).

[334] G. H. Booth, A. Grüneis, G. Kresse, and A. Alavi "Towards an exact description of electronic wavefunctions in real solids." *Nature* **493**, 365 (2013).

[335] A. Erba, S. Casassa, L. Maschio, and C. Pisani "DFT and local-MP2 periodic study of the structure and stability of two proton-ordered polymorphs of ice." *J. Phys. Chem. B* **113**, 2347 (2009).

[336] X. He, O. Sode, S. S. Xantheas, and S. Hirata "Second-order many-body perturbation study of ice Ih" *J. Chem. Phys.* **137**, 204505 (2012).

[337] G. A. Tribello, B. Slater, M. A. Zwijnenburg, and R. G. Bell "Isomorphism between ice and silica" *Phys. Chem. Chem. Phys.* **12**, 8597 (2010).

[338] C. Knight and S. Singer Programs to enumerate hydrogen bond topologies in ice, and generate graph invariants. Available from: `https://chemistry.osu.edu/~singer/GrEnum.html` (2014).

[339] S. J. Clark, M. D. Segall, C. J. Pickard, P. J. Hasnip, M. I. J. Probert, and K. Refson "First principles methods using CASTEP" *ZeitKristollagraphie* **220**, 567 (2005).

[340] K. Lee, E. D. Murray, L. Kong, B. I. Lundqvist, and D. C. Langreth "Higher-accuracy van der Waals density functional" *Phys. Rev. B* **82**, 081101 (2010).

[341] J. Klimeš, D. Bowler, and A. Michaelides "Van der Waals density functionals applied to solids" *Phys. Rev. B* **83**, 1 (2011).

[342] R. Sabatini, T. Gorni, and S. de Gironcoli "Nonlocal van der Waals density functional made simple and efficient" *Phys. Rev. B* **87**, 041108 (2013).

[343] Y. Zhao and D. G. Truhlar "The M06 suite of density functionals for main group thermochemistry, thermochemical kinetics, noncovalent interactions, excited states, and transition elements: two new functionals and systematic testing of four M06-class functionals and 12 other functionals" *Theor. Chem. Acc.* **120**, 215 (2007).

[344] F. Labat, C. Pouchan, C. Adamo, and G. E. Scuseria "Role of nonlocal exchange in molecular crystals: The case of two proton-ordered phases of ice." *J. Comput. Chem.* **32**, 2177 (2011).

[345] T. F. Whale, S. J. Clark, J. L. Finney, and C. G. Salzmann "DFT-assisted interpretation of the Raman spectra of hydrogen-ordered ice XV" *J. Raman Spectrosc.* **44**, 290 (2013).

[346] S. W. Rick "Simulations of proton order and disorder in ice Ih" *J. Chem. Phys.* **122**, 094504 (2005).

[347] M. P. Allen and D. J. Tildesley *Computer Simulation of Liquids* Clarendon Press, OUP (1989).

[348] M. Stengel, N. A. Spaldin, and D. Vanderbilt "Electric displacement as the fundamental variable in electronic-structure calculations" *Nature Phys.* **5**, 304 (2009).

[349] J. L. Aragones, L. G. Macdowell, and C. Vega "Dielectric Constant of Ices and Water: A Lesson about Water Interactions." *J. Phys. Chem. A* **115**, 5745 (2011).

[350] G. P. Johari and E. Whalley "Dielectric properties of ice VI at low temperatures" *J. Chem. Phys.* **64**, 4484 (1976).

[351] G. P. Johari and E. Whalley "Evidence for a very slow transformation in ice VI at low temperatures" *J. Chem. Phys.* **70**, 2094 (1979).

[352] S. W. de Leeuw, J. W. Perram, and E. R. Smith "Simulation of Electrostatic Systems in Periodic Boundary Conditions. I. Lattice Sums and Dielectric Constants" *Proc. R. Soc. London A* **373**, 27 (1980).

[353] G. Hummer, N. Grønbech-Jensen, and M. Neumann "Pressure calculation in polar and charged systems using Ewald summation: Results for the extended simple point charge model of water" *J. Chem. Phys.* **109**, 2791 (1998).

[354] S. M. Jackson and R. W. Whitworth "Evidence for ferroelectric ordering of ice Ih" *J. Chem. Phys.* **103**, 7647 (1995).

[355] G. M. Phillips and P. J. Taylor *Theory and Applications of Numerical Analysis* Academic Press, 2nd Edition (1996).

[356] Maple 13 Maplesoft, a division of Waterloo Maple Inc., Waterloo, Ontario.

[357] R. A. Kendall and H. A. Früchtl "The impact of the resolution of the identity approximate integral method on modern ab initio algorithm development" *Theor. Chem. Acc.* **97**, 158 (1997).

[358] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery *Numerical Recipes in FORTRAN;*

The Art of Scientific Computing Cambridge University Press, New York, NY, USA 2nd edition (1993).

[359] E. F. Valeev "A library for the evaluation of molecular integrals of many-body operators over Gaussian functions" http://libint.valeyev.net/ (2014).

[360] M. Dion, H. Rydberg, E. Schröder, D. C. Langreth, and B. I. Lundqvist "Van der Waals Density Functional for General Geometries" *Phys. Rev. Lett.* **92**, 246401 (2004).

[361] M. Dion, H. Rydberg, E. Schröder, D. C. Langreth, and B. I. Lundqvist "Erratum: Van der Waals Density Functional for General Geometries [Phys. Rev. Lett. 92, 246401 (2004)]" *Phys. Rev. Lett.* **95**, 109902 (2005).

[362] J. Klimeš, D. Bowler, and A. Michaelides "Chemical accuracy for the van der Waals density functional" *J. Phys.: Condens. Matter* **22**, 022201 (2010).